IRE Transactions

on INFORMATION THEORY

Volume IT-3 SEPTEMBER, 1957 Number 3

In This Issue

For complete Table of Contents, see page 171.

PUBLISHED BY THE
Professional Group on Information Theory

IRE Professional Group on Information Theory

The Professional Group on Information Theory is an organization, within the framework of the IRE, of members with principal professional interest in Information Theory. All members of the IRE are eligible for membership in the Group and will receive all Group publications upon payment of an annual assesment of $3.00.

IRE TRANSACTIONS® ON INFORMATION THEORY is published by the IRE for the Professional Group on Information Theory, at 1 East 79th Street, New York 21, N. Y. Responsibility for contents rests upon the authors and not upon the IRE, the Group, or its members. Price per copy: IRE-PGIT members, $2.20; IRE members, $3.30; nonmembers, $6.00.

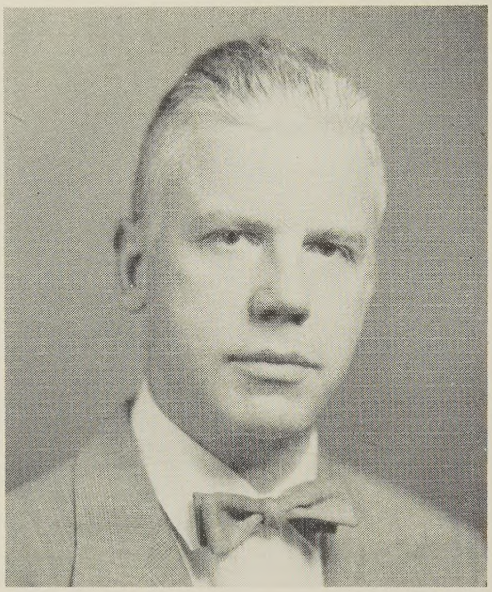
INFORMATION THEORY

IRE Transactions on Information Theory

Published Quarterly by the Professional Group on Information Theory

Volume IT-3 September, 1957 Number 3

TABLE OF CONTENTS



Brockway McMillan

Brockway McMillan (SM '54) was born on March 30, 1915 in Minneapolis, Minn. He attended Armour Institute, Chicago, Ill., from 1932 until 1934, when he entered Massachusetts Institute of Technology. There he received the B.S. degree in 1936 and the Ph.D. degree in 1939, both in mathematics. He was a part-time instructor in mathematics at M. I. T., during 1936-1939.

From 1939 to 1943, he was a Proctor Fellow, then Fine Instructor, and later research associate at Princeton University. In the latter capacity he served on a project studying the performance of optical and radar antiaircraft fire control systems. He was on active duty with the U. S. Naval Reserve from 1943 to 1946, in technical assignments at the U. S. Naval Proving Ground, Dahlgren, Va., and at the Los Alamos Laboratory.

Dr. McMillan joined the technical staff of the Bell Telephone Laboratories, Inc., in 1946 as a research mathematician. He has worked on the theory of physical realizability of multiterminal networks and the analysis of military weapon systems, but his principal interest has been the application of probability to communication problems. He holds several patents on amplifying and computing devices and is the author of nine technical papers. Since June, 1955, he has been Assistant Director of Systems Engineering I at the Bell Telephone Laboratories.

He is a vice-president of the Society for Industrial and Applied Mathematics, and a member of the American Mathematical Society, the Mathematics Association of America, the Institute of Mathematical Statistics, the Operations Research Society of America, and the American Association for the Advancement of Science.

Where Do We Stand?

BROCKWAY McMILLAN

One function of theory is to develop methods for solving actical problems. Another is to organize and to make mprehensible the domain of knowledge in which the eory operates. Those who apply theory place their own lues on these two functions, but all, in fact, use both. be solved, even the most practical of problems must stated in the terms of some underlying logical structure; solve it, even the most practical of engineers must mprehend this structure and understand the terms ereof.

The theory of communication offers the engineer a gical structure founded upon circuit theory and upon e theory of probability. Within this structure, it offers m methods for solving certain practical, and even npractical, problems. These are problems of detection. he theory recognizes the operation of detection as that decision or measurement in the presence of error. This exactly the problem to which the efforts of mathematical atisticians have been directed since early in the century. hus the tools of the statistician are available to the ngineer; he can now design detection systems, optimum y one of several criteria, for signals perturbed by one of veral different types of noise.

For the case in which the signals to be detected are me-series perturbed by added noise, statistical tools re particularly sharp; theory offers in this limited domain ethods of such elegance and power that they transcend ven the challenge of the practical problems for which ey were first conceived.

It is scarcely an understatement to say that problems detection are the only ones for which statistical theory ow gives the engineer specific methods of solution. heory has, however, gone farther than this in organizing is thought.

In the first place, the theory of detection itself can be, nd recently has been by Middleton and van Meter, nbedded in the more general theory of statistical decision, s developed by the late A. Wald and others. This latter neory, in turn, allows a very general measure of the erformance of a detector, derives general relations among ethods of detection, and in some cases delimits the omain within which the engineer must search to optimize s chosen measure of performance. All results at present late to the case in which the list of alternative decisions; e., the list of signals, is pregiven and fixed.

This decision theory of Wald provides a pattern within hich one can account for the undesirability of errors in etection by assigning almost arbitrarily an appropriate st to each error, and to indecision. The cost of error, wever, is not the only cost with which an engineer ust cope. He also has to reckon the cost of his equipment, either as a first cost, or as an operating cost measured either in money or more abstractly in terms of complexity, liability to failure, or sensitivity to deviations in components. Most such costs depend critically upon the accidents of available technology; it is too much to ask that a general theory account for them in detail. Yet there is one cost of this kind for which an effective accounting is possible.

When dealing with communication as a service—as distinct from communication as an adjunct to measurement or control—the engineer can fairly assume that revenue is a significant negative contributor to operating cost, and further, that the rate of revenue is proportional to the number of messages handled in a given time. Motivated by these assumptions, the work of Shannon comprehends in a single structure the engineer's view of the process of communication as a service.

Here appears the quantity of primary interest to such communication, namely, the communication rate in effective characters per unit time. Here appears the operation of modulation, in recognition that the list of alternative signals is not pregiven and fixed; indeed, the engineer designing a communication system has control, not merely over the mode of detection, but also over the actual form of the signals to be detected. Here also appears, and for the first time, the general concept of the received signal as a random entity merely distributed jointly with the transmitted signal.

Within this framework, Shannon's theory now exhibits a least upper bound to the achievable rates of communication. The bound is a quantity intrinsic to the medium of communication and its accompanying noise; to attain it exploits the full freedom of the engineer both to detect and to modulate optimally.

This theory also leaves room to discuss another abstractly measured element of (positive) cost: the complexity of terminals as measured by the delay required for the processes of modulation and detection. Little is now known, in general, about this latter quantity.

Most of the definitive statistical results in the theories just sketched describe the optima which can be achieved by the designer who knows exactly, and is able to exploit to its utmost, the statistical description of the universe within which his design must operate. Without denying the value of such results to the designer, and certainly without belittling the present accomplishments of theory, one can still urge consideration of other of the designer's problems.

In the first place, practically, the designer has only limited control over his medium. In principle, present theory has room for this phenomenon, since failure of the

implementation to realize exactly the designer's plan can be regarded as a form of noise. Unfortunately, we have at present no attractive or tractable general models for such noise; furthermore, except perhaps in extended systems of many similar tandem links, this kind of noise lacks the important simplifying property of ergodicity. There appears here to be a need both for new concepts and for new methods.

A fundamental weakness in present theory is the assumption of perfect knowledge on the part of the designer. In fact, design itself is a statistical decision. Rationally, it demands measurements of, rather than assumptions about, the statistical universes involved; this places it within the purview of Wald's theory o decision, or, alternatively, of the theory of games. T what extent the general results of these theories can b usefully adapted to problems of design remains an ope and interesting question.

Practically, because of the limitations, already noted of his basic data and of his available implementation the working engineer generally cannot now justify strivin for the ultimate optima to which present theory can lea him. Even in the absence of major improvements i theory, he could profit from the development of method which are conceptually simple and demonstrably in-sensitive to assumptions.

Detection of Fluctuating Pulsed Signals in the Presence of Noise*

PETER SWERLING†

Summary—This paper treats the detection of pulsed signals in ıe presence of receiver noise for the case of randomly fluctuating ignal strength. The system considered consists of a predetection tage, a square law envelope detector, and a linear postdetection ıtegrator. The main problem is the calculation of the probability ensity function of the output of the postdetection integrator. The nalysis is carried out for a large family of probability density ınctions of the signal fluctuations and for very general types of orrelation properties of the signal fluctuations. The effects of onuniform beam shape and of nonuniform weighting of pulses by ıe postdetection integrator are also taken into account. The function 'hich is actually evaluated is the Laplace transform of the prob- bility density function of the integrator output. In many of the ases treated, the resulting Laplace transform has an inverse of nown form. In such cases the evaluation of the probability density ınction would require the computation of a finite number of onstants; in practice this would usually require the use of com- uting machinery, but would be perfectly feasible with presently vailable computing machinery.

AN EXTENSIVE treatment of detection theory for pulsed signals in noise, for the case where the signal amplitude is constant, was given by Iarcum.[1,2] This analysis has been extended by other ıvestigators[3-5] to some cases where the signal pulse mplitudes are randomly modulated. However, almost ll analyses to date have dealt with just two cases, insofar s the correlation properties of the signal fluctuations are oncerned: the signal amplitudes fluctuate independently rom pulse to pulse, or the signal amplitudes are constant luring the integration time of the receiver but are inde- endent from one integration period to the next. Fluctua- ions not conforming to one of these assumptions have een treated only in very special cases.[4] The purpose of his paper is to extend the analysis to much more general ases, insofar as the correlation properties of the fluctua- ions are concerned. The analysis is carried out for a arge family of probability density functions for the ignal fluctuations. Also, since no additional work is ıvolved, we treat the case where the postdetection ıtegrator forms a weighted sum of the input pulses.

We shall consider a receiver consisting of a predetection tage, a square law envelope detector, and a linear post-detection integrator. The receiver noise at the detector input is assumed to be additive Gaussian (with zero mean), completely correlated for times of the order of one pulse width and completely uncorrelated from one pulse to the next. The detector is assumed to be a square law envelope detector. For mathematical convenience, the detector output is assumed to be normalized as follows: detector output equals input envelope squared divided by twice the input mean square receiver noise voltage. This normalization was used by Marcum[1,2] and followed by Swerling;[3] it simplifies some of the formulas, but results in no actual loss of generality.

Denoting by v_i the (normalized) ith pulse emerging from the detector, we assume the postdetection integrator to form the following weighted sum:

$$y = \text{integrator output} = \sum_{i=1}^{N} \alpha_i v_i \tag{1}$$

where α_i are positive real numbers.

We assume that the detection procedure requires the integrator output y to exceed a threshold Y_b in order for detection of a signal to be announced. Here Y_b is a dimensionless quantity, because of the normalization of the detector output described above.

If $G(y)$ represents the probability density function (pdf) for y, and if $G_0(y)$ represents the pdf in the case where receiver noise only is present, then

$$\text{Probability of detection} = \int_{Y_b}^{\infty} G(y)\, dy, \tag{2}$$

$$\text{Probability of false alarm} = \int_{Y_b}^{\infty} G_0(y)\, dy. \tag{3}$$

In most applications, the probability of false alarm is set at some desired level and Y_b is calculated from (3), then probability of detection is calculated from (2).

In case $\alpha_i = 1$, all i, then $G_0(y)$ is[2]

$$G_0(y) = \frac{1}{(N-1)!} y^{N-1} e^{-y} \quad (\text{for } \alpha_i = 1, \text{ all } i). \tag{4}$$

In the general case, the Laplace transform of $G_0(y)$ is given by

$$\int_0^{\infty} e^{-py} G_0(y)\, dy = \prod_{i=1}^{N} \frac{1}{\alpha_i p + 1}. \tag{5}$$

$$(\text{Real part of } p \geqq 0.)$$

This holds, of course, independent of any assumptions as to the signal fluctuations.

In order to calculate $G(y)$ in the presence of signal, it is necessary to specify the statistical nature of the signal

* Manuscript received by the PGIT, January 18, 1957.

† Control Systems Lab., University of Illinois, Urbana, Ill.

[1] J. I. Marcum, "A Statistical Theory of Target Detection by 'ulsed Radar," The RAND Corp., Res. Memo. RM-754; December 1, 1947.

[2] J. I. Marcum, "A Statistical Theory of Target Detection by 'ulsed Radar: Mathematical Appendix," The RAND Corp., Res. Iemo. RM-753; July 1, 1948.

[3] P. Swerling, "Probability of Detection for Fluctuating Targets," ıe RAND Corp., Res. Memo. RM-1217; March 17, 1954.

[4] M. Schwartz, "Effects of signal fluctuation in the detection of ulsed signals in noise," IRE Trans., vol. IT-2, pp. 66-71; June, 956.

[5] E. L. Kaplan, "Signal detection studies, with applications," Bell Sys. Tech. J., vol. 34; March, 1955.

fluctuations. We define x_i = ratio, at the detector input, of the signal power for the ith pulse to the mean receiver noise power.

We assume that x_i is of the following form:

$$x_i = \sum_{k=1}^{L} u_{k,i}^2 \qquad (i = 1, \cdots, N) \tag{6}$$

where L is a positive integer, and $u_{k,i}$ are Gaussian random variables with zero mean. (The x_i are also assumed to be statistically independent of the receiver noise.)

Define the random vectors $\mathbf{U}^{(1)}, \cdots, \mathbf{U}^{(L)}$by

$$\mathbf{U}^{(k)} = (u_{k,1}, \cdots, u_{k,N}) \qquad (k = 1 \cdots, L). \tag{7}$$

We assume that the $\mathbf{U}^{(k)}$, $k = 1, \cdots, L$, are mutually statistically independent. Also, it is assumed that $\mathbf{U}^{(k)}$ has covariance matrix $(\phi_{ij}^{(k)})$ where, denoting expected values by a bar,

$$\phi_{ij}^{(k)} = \overline{u_{k,i}u_{k,j}} \qquad \begin{array}{l} k = 1, \cdots, L \\ i,j = 1, \cdots, N. \end{array} \tag{8}$$

In radar applications, the fluctuation in x_i is considered to be due largely to fluctuation in the scattering cross section of the target. To relate the above formulation to more familiar types of fluctuation, consider the case where $L = 2K$ and where, for each i, $\overline{u_{1,i}^2} = \overline{u_{2,i}^2} = \cdots = \overline{u_{L,i}^2}$. Then it is easily verified that x_i has a pdf $w(x_i; \bar{x}_i)$ given by

$$w(x_i; \bar{x}_i) = \frac{1}{(K-1)!} \frac{K}{\bar{x}_i} \left(\frac{Kx_i}{\bar{x}_i}\right)^{K-1} \exp\left(\frac{-Kx_i}{\bar{x}_i}\right), \qquad x_i \geqq 0 \tag{9}$$

where exp ( ) stands for the exponential function.

Here $\bar{x}_i$ represents the average of x_i over the fluctuations. For $K = 1$, this reduces to the familiar exponential or Rayleigh fluctuation for the signal power.

Note that we do not require $\bar{x}_i$ to be the same for all i. This can be considered in radar applications to reflect the effect of beam shape.

For present purposes, no attempt will be made to derive the above formulation of the signal fluctuation from physical considerations. Its justification is simply that from it one can construct a wide variety of pdf's and correlation properties for the signal fluctuation.

We shall now compute the Laplace transform $C(p)$ of the pdf $G(y)$ of the integrator output y. This is defined as

$$C(p) = \int_0^\infty e^{-py} G(y)\, dy \tag{10}$$

where p is a complex number with nonnegative real part. Since $G(y)$ vanishes for $y < 0$, we are able to deal with the Laplace transform, integrating only from zero to infinity.

If we consider the conditional pdf for y, for definite values of $x_1, \cdots, x_N$, it is well known[2] that the Laplace transform $C(p|x_1, \cdots, x_N)$ of this conditional pdf is

$$C(p|x_1, \cdots, x_N) = \prod_{i=1}^{N} \frac{1}{1+\alpha_i p} \exp\left[\frac{-px_i\alpha_i}{1+\alpha_i p}\right]. \tag{11}$$

In view of (6), this can be written

$$C(p|x_1, \cdots, x_N) = \prod_{i=1}^{N} \frac{1}{1+\alpha_i p} \exp\left[\frac{-\alpha_i p \sum_{k=1}^{N} u_{k,i}^2}{1+\alpha_i p}\right]. \tag{12}$$

$C(p)$ is simply equal to this expression averaged over the probability distribution of $\mathbf{U}^{(1)}, \cdots, \mathbf{U}^{(L)}$. Because of the mutual independence of the $\mathbf{U}^{(k)}$, one may write

$$C(p) = \left[\prod_{i=1}^{N} \frac{1}{1+\alpha_i p}\right] \int \exp\left[-\sum_{i=1}^{N} \frac{\alpha_i p}{1+\alpha_i p} \cdot (u_{1,i}^2 + \cdots + u_{L,i}^2)\right] \prod_{k=1}^{L} dP(\mathbf{U}^{(k)}). \tag{13}$$

Now, assuming nonsingularity of $(\phi_{ij}^{(k)})$ for each k,

$$dP(\mathbf{U}^{(k)}) = \frac{1}{(2\pi)^{N/2}\Delta_k^{\frac{1}{2}}} \cdot \exp\left[-\tfrac{1}{2} \sum_{i,j=1}^{N} \xi_{ij}^{(k)} u_{k,i} u_{k,j}\right] d\mathbf{U}^{(k)} \tag{14}$$

where

$$\Delta_k = \text{determinant } (\phi_{ij}^{(k)}),$$
$$(\xi_{ij}^{(k)}) = \text{matrix inverse of } (\phi_{ij}^{(k)}).$$

Thus,

$$C(p) = \left[\prod_{i=1}^{N} \frac{1}{1+\alpha_i p}\right] \cdot \left[\prod_{k=1}^{L} E_k\right] \tag{15}$$

where

$$E_k = \frac{1}{(2\pi)^{N/2}\Delta_k^{\frac{1}{2}}} \int \exp\left\{-\tfrac{1}{2} \sum_{i,j=1}^{N} \xi_{ij}^{(k)} u_{k,i} u_{k,j} - \sum_{i=1}^{N} \frac{\alpha_i p}{1+\alpha_i p} u_{k,i}^2\right\} d\mathbf{U}^{(k)}. \tag{16}$$

Now define

$$\begin{aligned} \xi_{ij}^{(k)}(p) &= \xi_{ij}^{(k)}, & i \neq j \\ &= \xi_{ii}^{(k)} + \frac{2\alpha_i p}{1+\alpha_i p}, & i = j \end{aligned} \tag{17}$$

and

$$\Gamma_k(p) = \text{determinant } (\xi_{ij}^{(k)}(p)) \cdot \tag{18}$$

$\left(\text{Thus } \Gamma_k(0) = \dfrac{1}{\Delta_k}.\right)$

Then

$$E_k = \left[\frac{\Gamma_k(0)}{\Gamma_k(p)}\right]^{\frac{1}{2}} \left(\frac{[\Gamma_k(p)]^{\frac{1}{2}}}{(2\pi)^{N/2}} \int \exp\left\{-\tfrac{1}{2} \sum_{i,j=1}^{N} \xi_{ij}^{(k)}(p) u_{k,i} u_{k,j}\right\} d\mathbf{U}^{(k)}\right). \tag{19}$$

Now suppose for the moment that p is a nonnegative real number. Since the right-hand factor is then just the integral of a probability density and hence equal to unity,

$$E_k = \left[\frac{\Gamma_k(0)}{\Gamma_k(p)}\right]^{\frac{1}{2}}. \tag{20}$$

Thus, provided $(\phi_{ij}^{(k)})$ is nonsingular with matrix inverses $(\xi_{ij}^{(k)})$,

$$C(p) = \left\{\prod_{i=1}^{N} \frac{1}{1+\alpha_i p}\right\}\left\{\prod_{k=1}^{L}\left[\frac{\Gamma_k(0)}{\Gamma_k(p)}\right]^{\frac{1}{2}}\right\} \tag{21}$$

where $\Gamma_k(p)$ is defined by (17) and (18). This has been derived for p, a nonnegative real number, but by analytic continuation it clearly holds for all p in the closed right half plane. The only interesting singular case is that of complete correlation of the x_i, which can be treated as a separate case.[3-5]

It is interesting to consider some special cases. For example, suppose $\alpha_i = 1$, all i. Define

$$\begin{aligned} \lambda_i^{(k)}(p) &= i\text{th} \quad \text{eigenvalue of} \quad (\xi_{ij}^{(k)}(p)) \\ \lambda_i^{(k)}(0) &= \lambda_i^{(k)} = i\text{th} \quad \text{eigenvalue of} \quad (\xi_{ij}^{(k)}). \end{aligned} \tag{22}$$

Then, if $\alpha_i = 1$, all i, $\lambda_i^{(k)}(p) = \lambda_i^{(k)} + \dfrac{2p}{1+p}$, and

$$\begin{aligned} \Gamma_k(p) &= \prod_{i=1}^{N}\left(\lambda_i^{(k)} + \frac{2p}{1+p}\right), \\ \Gamma_k(0) &= \prod_{i=1}^{N}\lambda_i^{(k)}. \end{aligned} \tag{23}$$

Putting (23) into (21), for the case $\alpha_i = 1$, all i,

$$C(p) = \frac{1}{(1+p)^N}\prod_{k=1}^{L}\prod_{i=1}^{N}\left[1 + \frac{2p}{(1+p)\lambda_i^{(k)}}\right]^{-\frac{1}{2}}. \tag{24}$$

If the covariance matrices are all equal, $\phi_{ij}^{(k)} = \phi_{ij}$, all k, then $\lambda_i^{(k)} = \lambda_i$, all k, and

$$C(p) = \frac{1}{(1+p)^N}\prod_{i=1}^{N}\left[1 + \frac{2p}{(1+p)\lambda_i}\right]^{-L/2}. \tag{25}$$

Specializing still further, suppose that

$$\begin{cases} L = 2K \\ (\phi_{ij}^{(k)}) = (\phi_{ij}), \quad \text{all} \quad k \\ \alpha_i = 1, \quad \text{all} \quad i \\ \overline{u_{k,i}^2} = \phi_{ii} = \sigma^2, \quad \text{all} \quad i \quad \text{and} \quad k. \end{cases} \tag{26}$$

In these cases, x_i is distributed according to the pdf (9) and

$$\bar{x}_i = \bar{x} = 2K\sigma^2, \quad \text{all} \quad i. \tag{27}$$

The assumptions $\alpha_i = 1$, all i, and $\bar{x}_i = 2K\sigma^2$, all i, amount in radar applications to assuming a uniform beam, and uniform weighting of pulses by the postdetection integrator—assumptions almost invariably made in probability of detection calculations. However, even the assumptions listed in (26) allow a large degree of freedom in the choice of correlation properties and pdf's for the signal fluctuations.

Now define

$$\mu_i = i\text{th} \quad \text{eigenvalue of} \quad \left(\frac{\phi_{ij}}{\sigma^2}\right) = \frac{1}{\sigma^2\lambda_i}. \tag{28}$$

Using (25), (27), and (28), we obtain under the conditions listed in (26)

$$C(p) = (1+p)^{N(K-1)}\prod_{i=1}^{N}\frac{1}{\left[1 + p\left(1 + \dfrac{\bar{x}\mu_i}{K}\right)\right]^K} \tag{29}$$

where $\bar{x}$ is defined in (27), K in (26), and μ_i in (28). This can also be written

$$C(p) = \frac{1}{(1+p)^N}\prod_{i=1}^{N}\left[1 + \frac{\mu_i\bar{x}p}{K(1+p)}\right]^{-K}. \tag{30}$$

Special cases are

$$\begin{aligned} K = 1: C(p) &= \prod_{i=1}^{N}\frac{1}{1 + p(1+\mu_i\bar{x})} \\ K = 2: C(p) &= (1+p)^N\prod_{i=1}^{N}\frac{1}{\left[1 + p\left(1 + \dfrac{\mu_i\bar{x}}{2}\right)\right]^2}. \end{aligned} \tag{31}$$

These formulas (31) reduce to the correct formulas[3] when the signal strength fluctuates independently from pulse to pulse, in which case $\mu_i = 1$, all i. Also, they reduce to the correct formulas[3] in the case where the signal strength is constant for the N pulses $v_1, \cdots, v_N$; in this case, $\mu_i = 0$, $i = 1, \cdots, N-1$, and $\mu_N = N$. The correct formulas are obtained even though complete correlation leads to a singular covariance matrix (ϕ_{ij}); this indicates that the validity of (29) extends to some cases where (ϕ_{ij}) is singular.

It is interesting to note what happens to $C(p)$ in (30) when $K \to \infty$. Since, for $K \to \infty$, the standard deviation of the fluctuations about the mean value goes to zero, one would expect $C(p)$ to approach the form it would take for nonfluctuating signals. This is indeed what happens: from (30), as $K \to \infty$,

$$C(p) \to \frac{1}{(1+p)^N}\exp\left[\frac{-\bar{x}p}{1+p}\sum_{i=1}^{N}\mu_i\right].$$

But

$$\sum_{i=1}^{N}\mu_i = \text{trace of} \quad \left(\frac{\phi_{ij}}{\sigma^2}\right) = N,$$

so

$$C(p) \to \frac{1}{(1+p)^N}\exp\left[\frac{-N\bar{x}p}{1+p}\right]$$

(compare Marcum[2]).

One can use the expansion

$$\left(1 + \frac{a}{K}\right)^{-K} = \exp\left[-a + \frac{a^2}{2K} - \frac{a^3}{3K^2} + \cdots\right],$$

valid for $|\, a/K \,| < 1$, to obtain the following expression for $C(p)$:

$$C(p) = \frac{1}{(1+p)^N} \exp\left\{\frac{-N\bar{x}p}{1+p} \sum_{j=0}^{\infty} A_j \left[\frac{\bar{x}p}{K(1+p)}\right]^j\right\} \quad (32)$$

where

$$A_j = \frac{(-1)^j}{j+1}\left[\frac{1}{N}\sum_{i=1}^{N} \mu_i^{j+1}\right].$$

Eq. (32) is valid for all $p \geqq 0$ if $\bar{x}\mu_i/K \leqq 1$, all i; otherwise it is valid for

$$p < \min_i \left[\frac{1}{\frac{\bar{x}\mu_i}{K} - 1}\right].$$

It is useful to note that, in the cases represented by the assumptions listed in (26) and, in fact, in more general cases, the Laplace transform $C(p)$ can be inverted in a straightforward, though possibly very tedious, manner. The simplest case is, of course, $K = 1$ [see (31)], in which case, if for example the μ_i are distinct, the pdf $G(y)$ is of the form

$$G(y) = \sum c_i \exp\left[\frac{-y}{1+\mu_i\bar{x}}\right]$$

where the c_i are readily determined. Since digital computer programs exist for finding the eigenvalues of $N \times N$ matrices up to fairly high values of N, the above formulas can be utilized for digital machine computation of probability of detection in a fairly straightforward way.

Fixed Memory Least Squares Filters Using Recursion Methods*

MARVIN BLUM†

Summary—Given a set of equally spaced measurements, it is possible to curve fit a "least squares" polynomial to the N observed data points and obtain estimates of the past, present, or future values of the data or its derivatives by appropriate manipulations of the curve fit.

This curve fitting can be accomplished by a linear weighting of the observed data over an interval $(n-1)\ T$. If the data is measured in real time such that a new data point is observed each T seconds, then the desired output (for example, the smooth or predicted value of the data) can be obtained by sliding these fixed number of weights such that the same weight always multiplies the data which is at a fixed lag with respect to the most recent data. Since these weights are zero for lags greater than n, they may be described as a fix-finite memory linear digital filter.

In calculating the desired output for each new sample one requires a machine which can store n coefficients, n data points and performs n multiplications and $n-1$ additions in at least T seconds. The coefficients do not change but the multiplications and additions must be performed each T seconds as a new data point is measured.

For large values of n, and small T, this may put a severe requirement on the real time solutions of the computer. This paper presents an alternate technique using recursion formulas to obtaining the same results as the n point weighting equation. The method has the advantage of requiring considerably less storage, multiplications and additions when $n \gg 1$ and the degree of the curve fitting polynomial (K) is small.

* Manuscript received by the PGIT, January 21, 1957.
† Convair Astronautics, San Diego, Calif.

Fixed Memory Least Squares Filters Using Recursion Methods

Statement of the Problem

LET $y(t) = P(t) + N(t)$ be the input to a digital filter, where $P(t)$ is a polynomial in t of degree K, and $N(t)$ is a white noise source. The data are sampled every T units in time. It is desired to estimate the L^{th} derivative of $P(t)$ evaluated at $t = (m + \alpha)T$. The estimate is to be made in real time at $t = mT$. The filter has a finite memory and operates only on the present data point and the $(n - 1)$ previous data points. These properties may be summarized by the following equations.

Let Z_m be the output of the filter at $t = mT$ and y_{u+m-n} be the input to the filter at $t = (u + m - n)T$ then

$$Z_m = \sum_{u=1}^{n} C_u y_{u+m-n}, \quad \text{for} \quad m = 0, \pm 1, \pm 2. \quad (1)$$

The coefficients C_u are also functions of K, α, and L. However, these parameters are presumed fixed for a particular problem, and are deleted from the notation, to simplify the equations. The coefficients C_u are chosen

the standard least squares sense: that is,

$$E^1(Z_m) = \frac{d^L}{dt^L} P(t) \mid t = (m + \alpha)T \tag{2a}$$

$$E(Z_m - E(Z_m))^2 = \text{a minimum.} \tag{2b}$$

It is assumed that $E(N(t)) = 0$ for all t, and that $N(t)$ a stationary and ergodic white noise signal.

Eq. (1) can be interpreted as follows: the output Z_m given by a running weighted average over the n most recent data points. The weights are independent of m and so the same coefficients always multiply the input which is at a fixed lag with respect to the most current data. As an example, let

$$n = 3, \quad K = 1, \quad m = 3, 4 \text{ and } 5.$$

$$Z_3 = C_1y_1 + C_2y_2 + C_3y_3, \quad (m = 3), \tag{3}$$

Then

$$Z_4 = C_1y_2 + C_2y_3 + C_3y_4, \quad (m = 4),$$

$$Z_5 = C_1y_3 + C_2y_4 + C_3y_5, \quad (m = 5).$$

The use of (1) for large values of n can be burdensome. It is the purpose of this paper to determine a recursion procedure for computing Z_m which will be easier to apply than (1) for large n and small K.

For the remainder of the discussion it will be assumed that $T = 1$ since the required formulas and tables are simplest to define for this case. This imposes no essential restriction on the generality of the solution since it is only necessary to multiply the coefficients C_u by the magnitude of (T^{-L}) to obtain the answer for any other value of T, not equal to one.

Orthogonal Polynomials

In the tables of orthogonal polynomials by Anderson and Hauseman[2] and Fischer and Yates[3] are shown a listing of the first five polynomials and some of their important properties. Let $\xi_{V,n}(u)$ be a polynomial in u of degree V, where u takes on the integral values $u = 1, 2, 3, \cdots n$. Then the basic property of these polynomials is that

$$\sum_{u=1}^{n} \xi^2_{V,n}(u) = S(V, n) \tag{4}$$

and

$$\sum_{u=1}^{n} \xi_{V,n}(u)\xi_{W,n}(u) = 0, \; V \neq W.$$

The tabulated values[2,3] are for a function $\xi'_{V,n}(u)$ which is obtained from $\xi_{V,n}(u)$ by the relation

$$\xi'_{V,n}(u) = \lambda_{V,n}\xi_{V,n}(u). \tag{5}$$

The factor $\lambda_{V,n}$ appears in the solution for C_u in both the numerator and denominator as $\lambda^2_{V,n}$ and thus cancels. It is more convenient analytically to use $\xi_{V,n}(u)$ for arriving at a solution; however, once a solution has been obtained, the tabulated values for $\xi'_{V,n}(u)$ can be used for computational purposes.

It is shown[4] that if $P(u)$ (the input polynomial) is represented by

$$P(u) = \sum_{V=0}^{K} a_V\xi_{V,n}(u), \tag{6}$$

then the coefficients C_u are given by,

$$C_u = \sum_{V=L}^{K} \frac{\xi_{V,n}(u)\xi^{(L)}_{V,n}(n + \alpha)}{S(V,n)} \tag{7}$$

where

$$\xi^{(L)}_{V,n}(n + \alpha) \equiv \frac{d^L}{du^L} \xi_{V,n}(u) \mid_{u=n+\alpha}, \tag{8}$$

and u is treated as a continuous variable with respect to the differentiation.

Note that C_u is a polynomial of degree K in u. And since the coefficient u^K in $\xi_{K,n}(u)$ is unity, then

$$^5\,(D_{u-1})^K C_u = \frac{K!\xi^{(L)}_{V,n}(n + \alpha)}{S(V, n)}, \qquad u = 1, 2, 3, \cdots, n - K \tag{9}$$

where D^V_u ( ) is the difference operator which advances the index u, by V, of the bracketed quantity. Thus

$$D^V_u(C_u) = C_{u+V} \quad \text{or} \quad D^V_m(y_{u+m-n}) = y_{u+m-n+V}. \tag{10}$$

Eq. (9) is the symbolic representation of the operation, "K^{th} difference" and states that since C_u is a polynomial of degree K in u, the K^{th} difference of C_u with respect to u is a constant as given by (9).

Let the K^{th} difference with respect to m be taken of both sides of (1). Then

$$(D_m - 1)^K Z_m = (D_m - 1)^K \sum_{u=1}^{n} C_u y_{u+m-n},$$

or

$$(D_m - 1)^K Z_m = \sum_{u=1}^{n} C_u(D_m - 1)^K y_{u+m-n}. \tag{11}$$

Expanding $(D_m - 1)^K$, and noting (10), one obtains

$$(D_m - 1)^K Z_m = \sum_{u=1}^{n}\sum_{V=0}^{K} C_u(-1)^{K-V}\binom{K}{V} y_{u+m-n+V}, \tag{12}$$

[1] The operator $E(V)$ is the ensemble average of, or expected value of, the quantity (V).

[2] R. L. Anderson and E. E. Hauseman, "Tables of Orthogonal Polynomial Values Extended to $N=104$," Res. Bull. 297, Iowa State College, Ames, Ia.; April, 1942.

[3] R. A. Fischer and E. Yates, "Statistical Tables for Biological Agriculture and Medical Research," Oliver and Boyd, Ltd., Edinburgh and London; 1925.

[4] M. Blum, "An extension of the minimum mean square prediction theory for sampled input data," IRE Trans., Vol. IT-2, pp. 176–184; September, 1956.

[5] The operation $(D_{u-1})^K$ is the equivalent of the K^{th} difference on C_u.

where $\binom{K}{V}$ is the binomial coefficient $\frac{K!}{(K-V)!V!}$.

Let $u + V = L$, $L = 1, 2, 3, \cdots K$. Then the coefficient of y_{m-n+L} is defined as b_L, where

$$b_L = \sum_{u=1}^{L} C_u(-1)^{K+u-L}\binom{K}{L-u}. \tag{13}$$

Let $K + 1 \leqq u + V \leqq n$, then applying (9), the coefficient of $y_{m-n+K+1}$ to y_m is given by

$$B = \frac{K!(-1)^K \xi_K^{(L)}(n+\alpha)}{S(K, n)}. \tag{14}$$

Finally let $n + 1 \leqq u + V \leqq n + K$, then the coefficient of y_{m+d} ($d = 1, 2, 3, \cdots K$,) is given by δ_d, where

$$\delta_d = \sum_{V=d}^{K} C_{d+n-V}(-1)^{K-V}\binom{K}{V}. \tag{15}$$

Combining (12)-(15) one obtains

$$Z_{m+K} = -\sum_{V=0}^{K-1} Z_{m+V}(-1)^{K-V}\binom{K}{V} + \sum_{L=1}^{K} b_L y_{m-n+L} + \sum_{d=1}^{K} \delta_d y_{m+d} + B\sum_{j=1}^{n-K} y_{m-n+K+j}. \tag{16}$$

Let

$$\phi_m = \sum_{j=1}^{n-K} y_{m-n+K+j}, \tag{17}$$

then $\phi_m = y_m - y_{m-n+K} + \phi_{m-1}$, so that (16) may be written

$$Z_{m+K} = -\sum_{V=0}^{K-1} Z_{m+V}(-1)^{K+V}\binom{K}{V} + \sum_{L=1}^{K} [b_L y_{m-n+L} + \delta_L y_{m+L}] + B\phi_m. \tag{18}$$

Numerical Examples

Let the parameters be as follows:
$K = 2$ (quadratic input),
$\alpha = 0$, $L = 0$ (desired output is the smooth value of input), *e.g.*, $E(Z_m) = y_m$,
$n = 5$ (finite memory equal to $4T$), memory span over 5 most recent data points.
Then (7) becomes,

$$C_u = \sum_{V=0}^{2} \frac{\xi'_{V,5}(u)\xi'_{V,5}(5)}{S'(V, 5)},$$

or

$$C_u = 1/5 + \frac{\xi'_{1,5}(u)\xi'_{1,5}(5)}{S'(1, 5)} + \frac{\xi'_{2,5}(u)\xi'_{2,5}(5)}{S'(2, 5)}. \tag{19}$$

Using the tables of Anderson and Hauseman[6] one obtains the following:

$V \rightarrow$	1	2	
u	$\xi_1'(u)$	$\xi_2'(u)$	Cu
1	−2	+2	0.085714
2	−1	−1	−0.142857
3	0	−2	−0.085714
4	+1	−1	0.257143
5	+2	+2	0.885714
$S'(V,5)$	10	14	
$\lambda_{V,5}$	1	1	

Note

$$\xi_0'(u) = 1,$$

$$S'(0, n) = 1/n$$

and

$$\lambda_{0,n} = 1.$$

The output Z_m is given by

$$Z_m = \sum_{u=1}^{5} C_u y_{u+m-5}. \tag{20}$$

By the choice of parameters,

$$E(Z_m) = P(mT) \tag{21}$$

where $P(t)$ is a quadratic in t. Eq. (18) becomes,

$$Z_{m+2} = (2Z_{m+1} - Z_m + b_1 y_{m-4} + b_2 y_{m-3} + \delta_1 y_{m+1} + \delta_2 y_{m+2} + B\phi_m) \tag{22}$$

where

$$\phi_m = y_m - y_{m-3} + \phi_{m-1}, \tag{23}$$

and

$$\delta_d = \sum_{V=d}^{K} C_{d+n-V}(-1)^{K-V}\binom{K}{V}. \tag{24}$$

Therefore,

$$\delta_1 = C_4 - 2C_5 = -1.514285$$

$$\delta_2 = C_5 \qquad = 0.885714$$

and

$$b_L = \sum_{u=1}^{L} C_u(-1)^{K+u-L}\binom{K}{L-u}, \tag{25}$$

therefore,

$$b_1 = C_1 \qquad = 0.085714,$$

$$b_2 = C_2 - 2C_1 = -0.314285.$$

Finally from (14) one obtains,

$$B = \frac{K!(-1)^K \xi_{K.n}^{(L)}(n+\alpha)}{S(K, n)} = \frac{2!\lambda_{2,5}\xi'_{2,5}(5)}{S'(2, 5)}. \tag{26}$$

From table $n = 5$[7] one obtains,

[6] Anderson and Hauseman, *op. cit.*, p. 610.

[7] *Ibid.*

S' (2,5) $= 14$ directly below the table labeled ξ_2' and
λ (2,5) $= 1$, directly below S' (2,5), so that
$B = 0.285714$.

tarting the Recursion Process

After computing the constants as described above, ne still cannot start the computations of Z_{m+2}, [see (22)] nce one always requires two previous smoothed values $_{m+1}$, and Z_m. One may use (20) to compute Z_m and also $_{m+1}$ [by advancing m by 1 in (20)] but this would involve ost of the labor which this method seeks to avoid. An lternate procedure is to assume that $y_x = 0$ for $x < 0$ nd to start the process with m a large enough negative umber so that on the basis of the assumption, it is known nat Z_{m+1} and Z_m are zero. As an example in Table I one as computed, the value Z_{m+2} for $m = -3$ to $m = 5$ ssuming an input of the form,

$$y_x = x^2 + 0.5, \qquad x = 0, 1, 2, \cdots, \qquad (27)$$
$$= 0, \qquad x < 0.$$

n this case $N(t)$ has been taken as zero.

TABLE I

m	Z_{m+2}	y_{m+2}	Error in Z_{m+2}	ϕ_m
-3	0.	0.	0.	0.
-2	0.442857	0.5	0.057143	0.
-1	1.45714	1.5	-0.042858	0.
0	4.32857	4.5	-0.171431	0.5
1	9.37142	9.5	-0.128576	2.0
2	16.4999	16.5	0.	6.5
3	25.4999	25.5	0.	15.5
4	36.5000	36.5	0.	30.5
5	49.5000	49.5	0.	51.5

Note that the error due to the discontinuity at the rigin ($x = 0$) disappears as soon as n samples are obtained, ncluding the origin. This is due to the finite memory haracteristic of the filter. The zero values previous to the nemory span do not affect the results and one can coninue the calculations as if they were not introduced at ll. The particular example chosen was poor with respect o demonstrating the virtues of the recursion formula ecause n was so small. To demonstrate that the comolexity of the method is determined by K (the degree of he polynomial) and not by n, another example will be onsidered.

Let the parameters of the problem be taken as

$n = 10$ (memory span 10 samples),
$\left.\begin{matrix}\alpha = 1\\ L = 1\end{matrix}\right\}$ (prediction of derivative of the input at the next sample point: $E(Z_{m+2}) = P'((m+3)T)$),
$K = 2$ (quadratic input $P(t) = a_0 + a_1 t + a_2 t^2$).

Then from (7),

$$C_u = \sum_{V=1}^{2} \frac{\xi'_{V,10}(u)\xi'^{(1)}_{V,10}(11)}{S'(V, 10)}.$$

From Anderson and Hauseman[8]

$$\xi'_{1,10}(u) = \lambda_{1,10}(u - 11/2),$$

$$\frac{d}{du}\xi'_{1,10}(u)\,|_{u=11} = \lambda_{1,10} = 2,^{6}$$

$$\xi'_{2,10}(u) = \lambda_{2,10}((u - 11/2)^2 - 99/12),$$

$$\frac{d}{du}\xi'_{2,10}(u)\,|_{u=11} = \lambda_{2,10}\cdot(11) = 11/2.^{6}$$

One requires the formula,

$$Z_{m+2} = 2Z_{m+1} - Z_m + \delta_1 y_{m+1} + \delta_2 y_{m+2} + b_1 y_{m-9} + b_2 y_{m-8} + B\phi_m,$$

where

$$b_1 = C_1 = 0.195455,$$
$$b_2 = C_2 - 2C_1 = 0.35001,$$
$$\delta_1 = C_9 - 2C_{10} = -0.483333,$$
$$\delta_2 = C_{10} = 0.304545,$$

and

$$B = \frac{2!\lambda_{2,10}\xi'^{(1)}_{2,10}(11)}{S'(2, 10)} = 0.041667.$$

This is no more complex a solution than when $n = 5$. As K increases by one, the number of constants δ and b increase by one. The number of previous values of Z_m increases by one also. The complexity of the solution increases linearly with increasing K. Table II represents a comparison of the computing machine requirements for the weighed average technique, (1), and the recursion formula, (18).

TABLE II
COMPARISON OF COMPUTING MACHINE REQUIREMENTS FOR WEIGHTED AVERAGE FILTERING AND RECURSION FILTERING

n = number of samples
K = degree of polynomial passed without error by filter

Requirements	Weighted Average	Recursion Formula
Storage	n coefficients n: data points Total: $2n$	$2K + [K/2]$* coefficients $K + 1$ intermediate results $n + K$ data points Total $n + 4K + 1 + [K/2]$
Multiplication	n	$2K + 1 + [K/2]$
Additions	$n - 1$	$2(K + 1) + [K/2]$

*$[K/2]$ = smallest integer in the fraction, *e.g.*, $[1\frac{1}{2}] = 1$.

Again one may begin the solution by assuming,

$$y_x = 0,\ x < 0$$

and say, for instance,

[8] *Ibid.*, p. 598.

$y_x = x^2 + 0.5$ for $x \geq 0$ and let m

$$= -3, -2, -1, 0, 1, 2, 3.$$

One may verify by direct calculation that the transient error will exist for $m = -2$ to $m = 6$.

For $m \geq 7$, *e.g.*, when at least 10 data points from the curve, $y_x = x^2 + 0.5$, $(x \geq 0)$, have been used in the formula, the transient error due to negative x, will not effect the current answers.

Conclusion

A method has been presented for curve fitting to a polynomial in the least squares sense. The method has the advantage of using recursion formulas and thus avoiding longer calculation when the number of data points is large.

A method of starting the recursion process is suggested which has the advantage of simplicity of usage. A transient error is generated which however, disappears over the interval of one memory span of the filter.

Locally Stationary Random Processes*

RICHARD A. SILVERMAN†

Summary—A new kind of random process, the locally stationary random process, is defined, which includes the stationary random process as a special case. Numerous examples of locally stationary random processes are exhibited. By the generalized spectral density $\Psi(\omega, \omega')$ of a random process is meant the two-dimensional Fourier transform of the covariance of the process; as is well known, in the case of stationary processes, $\Psi(\omega, \omega')$ reduces to a positive mass distribution on the line $\omega = \omega'$ in the ω, ω' plane, a fact which is the gist of the familiar Wiener-Khintchine relations. In the case of locally stationary random processes, a relation is found between the covariance and the spectral density which constitutes a natural generalization of the Wiener-Khintchine relations.

Introduction

IN THIS PAPER we introduce the concept of a *locally stationary* random process, which is a natural generalization of the notion of a stationary random process. We shall show that there is a basic symmetry between the covariance and the spectral density of a locally stationary random process, and that the study of locally stationary random processes gives additional insight into the structure of stationary random processes. Before defining locally stationary random processes, we recall some relevant facts about random processes in general and stationary random processes in particular. For details, the reader is referred to the books of Loève,[1] Blanc-Lapierre and Fortet,[2] and Doob.[3]

Let $x(t)$ be a random process (in general complex), *i.e.*, a one-parameter family of random variables indexed by a real parameter t lying in some index set T, which we shall take to be a closed interval, usually the infinite real line R. We shall regard the parameter t as the time, and $x(t)$ as some physical noise process evolving randomly in time. We shall assume that the second moment of $x(t)$ exists for all $t \in T$, and that for all $t \in T$, the first moment of $x(t)$ is zero;[4] the latter assumption involves no loss of generality.[5] By the covariance of $x(t)$ is meant the function of two variables

$$\Gamma(t, t') = \langle x(t)x^*(t') \rangle, \qquad t, t' \in T, \tag{1}$$

defined on the product set $T \times T$; the angular parentheses denote the ensemble average, and the asterisk denotes the complex conjugate. It is apparent from (1) that $\Gamma(t, t')$ is *Hermitian*, *i.e.*, that $\Gamma(t, t') = \Gamma^*(t', t)$. A function $\Gamma(t, t')$ on $T \times T$ is said to be of *nonnegative definite type*, if for every finite subset T_n of T, and every complex-valued function $h(t)$

$$\sum_{t, t' \in T_n} \Gamma(t, t')h(t)h^*(t') \geq 0,$$

where t and t' separately range over all the points of T_n. A fundamental theorem states that *a function* $\Gamma(t, t')$ *on* $T \times T$ *is a covariance if and only if it is of nonnegative definite type.*[6] More specifically, if $\Gamma(t, t')$ is of nonnegative definite type, one can always construct a Gaussian random process which has $\Gamma(t, t')$ as its covariance. It should be noted that any positive constant is a covariance.

The random process $x(t)$, defined for all real t, is said to be *stationary* in the wide sense, or to have a *stationary covariance*, if its covariance is of the form

$$\Gamma(t, t') = \langle |x|^2 \rangle C(t - t'), \qquad t, t' \in R. \tag{2}$$

The quantity $\langle |x|^2 \rangle$ is the average instantaneous power of the process $x(t)$, and is independent of the time. We

* Manuscript received by the PGIT, January 25, 1957. The research reported in this article was sponsored in part by Contract No. DA49-170-sc-2253.

† New York University, Inst. of Math. Sci., New York, N. Y.

[1] M. Loève, "Probability Theory," D. Van Nostrand Co., Inc., New York, N. Y., ch. 10; 1955.

[2] A. Blanc-Lapierre and R. Fortet, "Théorie des Fonctions Aléatoires," Masson, Paris; 1953.

[3] J. L. Doob, "Stochastic Processes," John Wiley & Sons, Inc., New York, N. Y.; 1953.

[4] The symbol $\in$ means *is a member of*; the product set $T \times T$ is the set of all pairs (t, t') with $t, t' \in T$.

[5] M. Loève, *op. cit.*, p. 465.

[6] *Ibid.*, p. 466.

ave chosen the correlation function or stationary co-ariance[7] $C(\tau)$ to be *normalized*, by which we mean that $(0) = 1$. The fact that $\Gamma(t, t')$ is Hermitian implies that $(\tau) = C^*(-\tau)$. Suppose now that $C(\tau)$ is continuous ontinuity at the origin is sufficient) and that

$$\int_{-\infty}^{\infty} | C(\tau) | \, d\tau < \infty .$$

hen it follows[8] that there is a continuous *nonnegative* unction $\Phi(\omega)$, called the *spectral density* of the process (t), such that

$$\langle | x |^2 \rangle C(\tau) = \int_{-\infty}^{\infty} \exp (i\omega\tau) \Phi(\omega) \, d\omega, \tag{3}$$

nd

$$\Phi(\omega) = \frac{\langle | x |^2 \rangle}{2\pi} \int_{-\infty}^{\infty} \exp (-i\omega\tau) C(\tau) \, d\tau, \tag{4}$$

e., such that $\langle | x |^2 \rangle C(\tau)$ and $\Phi(\omega)$ are Fourier transform airs.[9] In particular, we have

$$\int_{-\infty}^{\infty} \Phi(\omega) \, d\omega = \langle | x |^2 \rangle .$$

qs. (3) and (4), together with the stated requirements n the functions $C(\tau)$ and $\Phi(\omega)$, constitute one form of he famous Wiener-Khintchine relations. It can also be hown[10,11] that the process $x(t)$ itself has the representation

$$x(t) = \int_{-\infty}^{\infty} \exp (i\omega t) \, d\hat{x}(\omega), \tag{5}$$

here $\hat{x}(\omega)$ is another random process, and where the ntegral is meant in the mean square sense.[12] The process (ω) is one of *orthogonal increments*,[13] which means in our ase that

$$\langle d\hat{x}(\omega) \, d\hat{x}^*(\omega') \rangle = \Phi(\omega) \, \delta(\omega - \omega') \, d\omega \, d\omega',$$

here $\delta(\omega - \omega')$ is the Dirac delta function, and $\Phi(\omega)$ is he spectral density of $x(t)$ defined by (4).

We need the following facts about the harmonic analysis f *nonstationary* random processes. Suppose that $x(t)$ is *armonizable*,[14] *i.e.*, that $x(t)$ can be written in the form (5), here $\hat{x}(\omega)$ has a covariance of bounded variation on $R \times R$. It follows that

$$\Gamma(t, t') = \left\langle \left[\int_{-\infty}^{\infty} \exp (i\omega t) \, d\hat{x}(\omega) \right] \cdot \left[\int_{-\infty}^{\infty} \exp (i\omega' t') \, d\hat{x}(\omega') \right]^* \right\rangle$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp [i(\omega t - \omega' t')] \langle d\hat{x}(\omega) \, d\hat{x}^*(\omega') \rangle, \tag{6}$$

a fact which is summarized by saying that $\Gamma(t, t')$ is *harmonizable*.[14] We shall assume that any random process with which we are concerned is harmonizable, that

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} | \Gamma(t, t') | \, dt \, dt' < \infty ,$$

except in the stationary case, and that

$$\langle d\hat{x}(\omega) \, d\hat{x}^*(\omega') \rangle = \Psi(\omega, \omega') \, d\omega \, d\omega',$$

where $\Psi(\omega, \omega')$ is a continuous function.[15] The function $\Psi(\omega, \omega')$ will be called the *generalized spectral density* of the process $x(t)$. Thus (6) becomes

$$\Gamma(t, t') = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp [i(\omega t - \omega' t')] \Psi(\omega, \omega') \, d\omega \, d\omega'. \tag{7}$$

In particular, we have

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Psi(\omega, \omega') \, d\omega \, d\omega' = \Gamma(0, 0).$$

The function $\Psi(\omega, \omega')$ is a covariance by construction, and obeys the inversion formula

$$\Psi(\omega, \omega') = \frac{1}{(2\pi)^2} \cdot \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp [-i(\omega t - \omega' t')] \Gamma(t, t') \, dt \, dt'. \tag{8}$$

Eqs. (7) and (8) are the natural generalizations of (3) and (4). As noted by Blanc-Lapierre and Fortet,[16] in the stationary case $\Psi(\omega, \omega')$ is a positive mass distribution concentrated on the line $\omega = \omega'$ in the ω, ω' plane. Indeed, any generalized spectral density corresponding to a continuous mass distribution on the line $\omega = \omega'$ can be written

$$\Psi(\omega, \omega') = \Phi(\omega) \, \delta(\omega - \omega') \tag{9}$$

where

$$\Phi(\omega) = \int_{-\infty}^{\infty} \Psi(\omega, \omega') \, d\omega' \geq 0.$$

It follows immediately, after substituting (9) in (7), that the covariances corresponding to distributions of the form (9) are stationary.

[7] The terms *correlation function* and *stationary covariance* will be sed interchangeably.

[8] J. L. Doob, *op. cit.*, p. 522.

[9] If the requirement that $C(\tau)$ be absolutely integrable is relaxed, 3) and (4) are replaced by more general expressions (see J. L. Doob, *p. cit.*, p. 519). This requirement on $C(\tau)$ corresponds to limiting he memory of the process $x(t)$ in a way which is quite reasonable hysically.

[10] M. Loève, *op. cit.*, p. 483.

[11] J. L. Doob, *op. cit.*, p. 527.

[12] M. Loève, *op. cit.*, p. 472.

[13] *Ibid.*, p. 479.

[14] *Ibid.*, p. 474.

[15] Again, these assumptions are not the most general mathematically, but are quite appropriate physically. In particular, $\Gamma(t, t')$ must be continuous and bounded. In the stationary case, $\Gamma(t, t')$ cannot be absolutely integrable in the two variables t and t'; instead we require that it be integrable in the difference variable $\tau = t - t'$. In the white noise case, we avoid infinite average instantaneous power by using the correlation function $\Delta(\tau)$ (see below), and corresponding to the discontinuity of $\Delta(\tau)$ at $\tau = 0$, we permit $\Gamma(t, t')$ to be discontinuous when $t = t'$. Similarly, in the stationary case, we permit $\Psi(\omega, \omega')$ to be discontinuous when $\omega = \omega'$, corresponding to the discontinuity at the origin of the Dirac delta function [see (9)].

[16] A. Blanc-Lapierre and R. Fortet, *op. cit.*, p. 382.

Locally Stationary Random Processes[17]

We shall say that the random process $x(t)$, defined for all real t, is *locally stationary* in the wide sense, or has a *locally stationary covariance*, if its covariance can be written as

$$\Gamma(t, t') = \left\langle \left| x\left(\frac{t+t'}{2}\right) \right|^2 \right\rangle C(t - t'), \qquad t, t' \in R, \quad (10)$$

where again $C(\tau)$ is a (normalized) stationary covariance. The symmetrized variable $(t + t')/2$ has been chosen because it makes $\Gamma(t, t')$ Hermitian and because of its suggestive meaning as the average or centroid of the points t and t'. Qualitatively, a locally stationary random process has a covariance equal to a stationary covariance multiplied by a sliding power factor; this factor continuously renormalizes the average instantaneous power to a representative local level, chosen as the power level at the centroid of the points t and t'. For simplicity, we denote the functions in the right side of (10) by Γ_1 and Γ_2, respectively. With this notation, (10) becomes

$$\Gamma(t, t') = \Gamma_1\left(\frac{t+t'}{2}\right)\Gamma_2(t - t'), \qquad t, t' \in R, \quad (11)$$

and we rephrase our previous definition as follows: *a locally stationary covariance is a covariance which can be written in the form* (11) *where* Γ_1 *is a nonnegative function, and* Γ_2 *is a stationary covariance.* We note at once that if Γ_1 is a positive constant, then (11) reduces to a stationary covariance. This shows that our definition of local stationarity has the desirable property of including stationarity as a special case. We shall now exhibit less trivial examples of locally stationary random processes.

We begin by considering the function $\Delta(t - t')$, which is 1 when $t = t'$ and 0 otherwise. This function is obviously of nonnegative definite type on $R \times R$, since if T_n is any finite set of real numbers and $h(t)$ is any complex-valued function, then

$$\sum_{t,t' \in T_n} \Delta(t - t')h(t)h^*(t') = \sum_{t \in T_n} | h(t) |^2 \geq 0.$$

$\Delta(t - t')$ is therefore a covariance, in particular a normalized stationary covariance. Moreover, the product

$$\Gamma_1\left(\frac{t+t'}{2}\right)\Delta(t - t'), \quad (12)$$

where Γ_1 is *any* nonnegative function (not necessarily a covariance), is also of nonnegative definite type and therefore a covariance, since

$$\sum_{t,t' \in T_n} \Gamma_1\left(\frac{t+t'}{2}\right)\Delta(t - t')h(t)h^*(t') = \sum_{t \in T_n} \Gamma_1(t) | h(t) |^2 \geq 0.$$

It follows that (12) is a locally stationary covariance. A process with a covariance of the form (12) will be called a *locally stationary white noise.* (The word *white* reflects the fact that $\Delta(\tau)$ is the symbolic Fourier transform of a constant in the sense described in the Appendix.)

The example just given shows that the product $\Gamma_1\Gamma_2$ can be a covariance without both Γ_1 and Γ_2 being covariances. However, if both Γ_1 and Γ_2 are covariances the product $\Gamma_1\Gamma_2$ is automatically a covariance.[18] It follows that the product of a nonnegative covariance of the form $\Gamma_1[(t + t')/2]$ with a stationary covariance $\Gamma_2(t - t')$ is a locally stationary covariance. Covariances of the form $\Gamma_1[(t + t')/2]$ have been studied by Loève,[19,20] who calls them *exponentially convex covariances.* As noted by Loève, any two-sided Laplace transform of a nonnegative function is an exponentially convex covariance. The proof is immediate, for suppose that

$$\Gamma_1(\tau) = \int_{-\infty}^{\infty} \exp(-\alpha\tau)p(\alpha)\,d\alpha, \qquad \tau \in T,$$

where $p(\alpha)$ is a nonnegative function. Then, for any finite subset T_n of T

$$\sum_{t,t' \in T_n} \Gamma_1\left(\frac{t+t'}{2}\right)h(t)h^*(t') = \int_{-\infty}^{\infty} \Big| \sum_{t \in T_n} \exp(-\alpha t/2)h(t) \Big|^2 p(\alpha)\,d\alpha \geq 0,$$

so that $\Gamma_1[(t + t'/2)]$ is of nonnegative definite type, and hence a covariance (on $T \times T$). Conversely, Loève notes that any continuous exponentially convex covariance $\Gamma_1(\tau)$, $\tau \in T$, can be represented as the two-sided Laplace transform of a nonnegative function in the absolutely continuous case. In particular, an exponentially convex covariance is nonnegative.

We can now exhibit a large class of locally stationary covariances by using the fact that the product of two covariances is a covariance. In fact, let $\Gamma_1(\tau)$ be any Laplace transform of a nonnegative function which converges for all $\tau \in R$, and let $\Gamma_2(\tau)$ be any correlation function. Then the product

$$\Gamma_1\left(\frac{t+t'}{2}\right)\Gamma_2(t - t'), \qquad t, t' \in R,$$

is a locally stationary covariance, which we shall call an *exponentially convex locally stationary covariance.* It can be seen at once that every exponentially convex covariance $\Gamma_1(\tau)$ defined for all $\tau \in R$ is exponentially unbounded either at $-\infty$ or $+\infty$, and consequently has no Fourier transform (except for the case of a positive constant which has a symbolic Fourier transform proportional to the Dirac delta function). It follows that the same is true of exponentially convex locally stationary covariances.

[17] In this connection, see the definition of *local homogeneity* given by A. N. Kolmogoroff, "The local structure of turbulence in incompressible viscous fluid for very large Reynolds number," *Doklady Akad. Nauk SSSR*, vol. 30, pp. 301–305; 1941.

[18] M. Loève, *op. cit.*, p. 468.

[19] M. Loève, "Fonctions aléatoires à décomposition orthogonale exponentielle," *Rev. Sci.*, vol. 84, pp. 159–162; 1946.

[20] M. Loève, "Fonctions Aléatoires du Second Ordre," in P. Lévy, "Processus Stochastiques et Mouvement Brownien," Gauthier-Villars, Paris; 1948.

owever, it is possible to convert suitable exponentially nvex locally stationary covariances into locally stationary covariances with Fourier transforms by the llowing simple device.

We begin by considering the equality

$$\exp[-a(t^2 + t'^2)] = \exp\left[-2a\left(\frac{t + t'}{2}\right)^2\right] \cdot \exp\left[-\frac{a}{2}(t - t')^2\right], \qquad a > 0. \tag{13}$$

he left side of (13) is obviously a covariance, since

$$\exp[-a(t^2 + t'^2)] = \exp(-at^2)\exp(-at'^2).$$

he first factor in the right side of (13) is a nonnegative unction which is not itself a covariance, since it is of the orm $\Gamma_1[(t + t')/2]$ and is not an exponentially convex ovariance. The second factor in the right side of (13) is stationary covariance, since the Fourier transform of xp $(-a\tau^2/2)$ is a nonnegative function. It follows that 13) is a locally stationary covariance, and in particular, ne which is not the product of two covariances.

Now let $x(t)$ be a random process with the exponentially onvex locally stationary covariance

$$\Gamma(t, t') = \Gamma_1\left(\frac{t + t'}{2}\right)\Gamma_2(t - t'),$$

nd let $y(t)$ be a new random process defined by

$$y(t) = \exp(-at^2)\, x(t), \qquad a > 0.$$

Using (13) we can write the covariance of $y(t)$ as

$$\Gamma_y(t, t') = \exp\left[-2a\left(\frac{t + t'}{2}\right)^2\right]\Gamma_1\left(\frac{t + t'}{2}\right) \cdot \exp\left[-\frac{a}{2}(t - t')^2\right]\Gamma_2(t - t'), \tag{14}$$

where Γ_1 is an exponentially convex covariance, and Γ_2 s a stationary covariance. The product of two nonnegative unctions is nonnegative, and the product of two stationary ovariances is a stationary covariance. It follows that the product of two locally stationary covariances is a locally tationary covariance. In particular, $\Gamma_y(t, t')$ is locally tationary. Suppose now that Γ_2 is absolutely integrable, nd that the function $p(\alpha)$ appearing in the definition of Γ_1 is such that the integral

$$\int_{-\infty}^{\infty} \exp(-2a\tau^2)\Gamma_1(\tau)\, d\tau = \sqrt{\frac{\pi}{2a}} \int_{-\infty}^{\infty} \exp(\alpha^2/8a)p(\alpha)\, d\alpha$$

s finite.[21] Then it follows that (14) is absolutely integrable nd has a Fourier transform. Thus, this device enables us o convert a large class of exponentially convex locally stationary covariances into locally stationary covariances with Fourier transforms.

[21] This is the case, for example, for any $a > 0$, if $p(\alpha)$ vanishes outside a finite interval.

It is apparent from the foregoing that well-behaved locally stationary covariances exist in abundance.

The Generalized Wiener-Khintchine Relations

We now prove an interesting generalization of the Wiener-Khintchine relations for the case of locally stationary random processes.

Theorem: *If $\Gamma(t, t')$ is a locally stationary covariance, then the generalized spectral density defined by (8) is also a locally stationary covariance.*[22]

Proof: Write (8) as

$$\Psi(\omega, \omega') = \frac{1}{(2\pi)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\left[-i\left(\frac{t + t'}{2}\right)(\omega - \omega') - i(t - t')\left(\frac{\omega + \omega'}{2}\right)\right]\Gamma(t, t')\, dt\, dt', \tag{15}$$

and introduce the new variables $u = (t + t')/2$ and $v = t - t'$. Substituting from (11), we see at once that the function $\Psi(\omega, \omega')$, which is itself a covariance, can be written

$$\Psi(\omega, \omega') = \Psi_1\left(\frac{\omega + \omega'}{2}\right)\Psi_2(\omega - \omega'), \tag{16}$$

where

$$\Psi_1(w) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(-ivw)\Gamma_2(v)\, dv, \tag{17}$$

$$\Psi_2(z) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(-iuz)\Gamma_1(u)\, du, \tag{18}$$

and

$$\Gamma_1(u) = \int_{-\infty}^{\infty} \exp(iuz)\Psi_2(z)\, dz, \tag{19}$$

$$\Gamma_2(v) = \int_{-\infty}^{\infty} \exp(ivw)\Psi_1(w)\, dw. \tag{20}$$

Since Γ_1 is a nonnegative function, Ψ_2 is a correlation function, and since Γ_2 is a continuous correlation function, Ψ_1 is a nonnegative function. It follows that $\Psi(\omega, \omega')$ is a locally stationary covariance, as asserted. We shall call (11), (16), and (17)-(20), together with the stated requirements on the functions involved, the *generalized Wiener-Khintchine relations*.

The case of stationary random processes is included in these relations by permitting $\Gamma_1(\tau)$ to be the positive constant $\Gamma_1(0) = \langle |x|^2 \rangle$, so that the covariance (11) reduces to (2), its form in the stationary case. We also admit, as is customary, the symbolic relations

$$1 = \int_{-\infty}^{\infty} \exp(i\xi\eta)\delta(\xi)\, d\xi, \qquad \delta(\xi) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(-i\xi\eta)\, d\eta, \tag{21}$$

[22] As remarked in the Introduction, we assume that $\Gamma(t, t')$ is continuous (except in the white noise case) and absolutely integrable in t and t' (except in the stationary case).

where $\delta(\xi)$ is the Dirac delta function which, as shown in the Appendix, can be regarded as an unnormalized correlation function. It follows that in the stationary case, the spectral density $\Psi(\omega, \omega')$ reduces to the special form [see (4)]

$$\Psi(\omega, \omega') = \Phi\left(\frac{\omega + \omega'}{2}\right)\delta(\omega - \omega') \equiv \Phi(\omega)\,\delta(\omega - \omega'), \tag{9}$$

where $\Phi(\omega)$ is a nonnegative function; as already noted, (9) is tantamount to the usual Wiener-Khintchine relations. The covariance (9) is obviously locally stationary and is, in fact, the covariance of a locally stationary white noise. (To see this, use the relation $\delta(\xi) = (1/\epsilon)\Delta(\xi)$ discussed in the Appendix.)

Other illustrations of the generalized Wiener-Khintchine relations are the following:

1). Let $\Gamma(t, t')$ be the locally stationary white noise

$$\Gamma(t, t') = \Gamma_1\left(\frac{t + t'}{2}\right)\Delta(t - t'),$$

where Γ_1 is any continuous and integrable nonnegative function. The corresponding spectral density is

$$\Psi(\omega, \omega') = \frac{\epsilon}{2\pi} R(\omega - \omega'), \tag{22}$$

where ϵ is the "infinitesimal positive constant" described in the Appendix, and where

$$R(\xi) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(-i\xi\eta)\Gamma_1(\eta)\,d\eta.$$

Since $R(\xi)$ is a stationary covariance, it is obvious that (22) is a locally stationary covariance.

2). Suppose $\Gamma(t, t')$ is the locally stationary covariance

$$\exp\left[-2a\left(\frac{t + t'}{2}\right)^2\right]\exp\left[-\frac{a}{2}(t - t')^2\right], \quad a > 0. \tag{13}$$

Then the generalized spectral density is

$$\Psi(\omega, \omega') = \frac{1}{4\pi a}\exp\left[-\frac{1}{2a}\left(\frac{\omega + \omega'}{2}\right)^2\right] \cdot \exp\left[-\frac{1}{8a}(\omega - \omega')^2\right], \tag{23}$$

which is immediately seen to be locally stationary since, except for a positive factor, it is of the form (13), with a replaced by $1/4a$.

In the case of locally stationary random processes, there is a manifest symmetry between the form of the covariance $\Gamma(t, t')$ and that of the spectral density $\Psi(\omega, \omega')$. The reason this symmetry is absent in the case of stationary random processes is apparent. The spectral density of a stationary random noise is itself stationary only when the noise is white. In general, the spectral density of a stationary noise is the covariance of a locally stationary white noise of the form (9), where the delta function appears as the Fourier transform of the *constant* average instantaneous power which characterizes a stationary process.

Conclusion

It is apparent that much remains to be done in the theory of locally stationary random processes. Perhaps the most important problem is to investigate further the class of noise processes which are either locally stationary or can be approximated in some appropriate fashion by locally stationary random processes. Heuristically speaking, if a process is such that its average instantaneous power is slowly varying with respect to its memory or correlation time, then it should be possible to approximate it by a locally stationary process. (A familiar example of such a process is the turbulent velocity field at large Reynolds number.[17]) Numerous other problems in the theory of locally stationary random processes are suggested by analogous problems in the theory of stationary processes.

The author wishes to acknowledge helpful suggestions by Dr. I. Kay and Professor J. E. Moyal.

Appendix

Some Properties of the Functions $\delta(\xi)$ and $\Delta(\xi)$

The Dirac delta function $\delta(\xi)$ can be regarded as the limit as $\epsilon \to 0$ of the function

$$\delta_\epsilon(\xi) = \frac{1}{\epsilon}\left(1 - \frac{|\xi|}{\epsilon}\right), \qquad |\xi| \le \epsilon > 0,$$

$$\delta_\epsilon(\xi) = 0, \qquad |\xi| > \epsilon.$$

Since

$$\delta_\epsilon(\xi) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(-i\xi\eta)\frac{\sin^2(\eta\epsilon/2)}{(\eta\epsilon/2)^2}\,d\eta,$$

$\delta_\epsilon(\xi)$ is the Fourier transform of a nonnegative function, and hence it is a correlation function. It follows that $\delta(\xi)$ is a correlation function, since it is the limit of a sequence of correlation functions.[18] (A simpler proof follows from the proportionality of $\delta(\xi)$ and $\Delta(\xi)$ described below, since $\Delta(\xi)$ is a correlation function, as previously noted.) Moreover, since

$$\lim_{\epsilon\to 0}\frac{\sin^2(\eta\epsilon/2)}{(\eta\epsilon/2)^2} = 1, \qquad \text{for all } \eta,$$

we have the symbolic relations

$$1 = \int_{-\infty}^{\infty} \exp(i\xi\eta)\,\delta(\xi)\,d\xi,$$

$$\delta(\xi) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(-i\xi\eta)\,d\eta. \tag{21}$$

The function $\Delta(\xi)$, defined to be 1 when $\xi = 0$ and 0 otherwise, can be regarded as the limit as $\epsilon \to 0$ of the function

$$\Delta_\epsilon(\xi) = 1 - \frac{|\xi|}{\epsilon}, \qquad |\xi| \le \epsilon > 0,$$

$$\Delta_\epsilon(\xi) = 0, \qquad |\xi| > \epsilon.$$

It will be observed that $\Delta_\epsilon(\xi) = \epsilon\delta_\epsilon(\xi)$. $\Delta(\xi)$ is a correlation function, since $\Delta(t - t')$ is obviously of nonnegative

definite type (or, alternatively, since it is the limit of a sequence of correlation functions). Symbolically, we can write

$$\Delta(\xi) = \epsilon\delta(\xi),$$

$$\delta(\xi) = \frac{1}{\epsilon}\Delta(\xi),$$

where ϵ is an "infinitesimal positive constant" which adjusts $\delta(0) = \infty$ to $\Delta(0) = 1$. The symbolic meaning of the relations

$$\epsilon = \int_{-\infty}^{\infty} \exp(i\xi\eta)\Delta(\xi)\, d\xi,$$
$$\Delta(\xi) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(-i\xi\eta)\epsilon\, d\eta, \tag{24}$$

obtained by formal multiplication of the relations (21) by ϵ is apparent. $\Delta(\xi)$ can be regarded as the correlation function of stationary white noise with unit average instantaneous power, just as $\delta(\xi)$ can be regarded as the correlation function of stationary white noise with infinite average instantaneous power.

The Solution of a Homogeneous Wiener-Hopf Integral Equation Occurring in the Expansion of Second-Order Stationary Random Functions*

D. C. YOULA†

Summary—In many of the applications of probability theory to problems of estimation and detection of random functions an eigenvalue integral equation of the type

$$\phi(x) = \lambda \int_0^T K(x - y)\phi(y)\, dy, \qquad 0 \le x \le T,$$

is encountered where $K(x)$ represents the covariance function of a continuous stationary second-order process possessing an absolutely continuous spectral density.

In this paper an explicit operational solution is given for the eigenvalues and eigenfunctions in the special but practical case when the Fourier transform of $K(x)$ is a rational function of ω^2, *i.e.*,

$$K(x) \doteqdot G(s^2) = \frac{N(s^2)}{D(s^2)}, \qquad s = i\omega,$$

in which $N(s^2)$ and $D(s^2)$ are polynomials in s^2.

It is easy to show by elementary methods that the solutions are of the form[1]

$$\phi(x) = \sum_r C_r e^{-\alpha_r x} \cos(\beta_r x + \gamma_r),$$

the constants C_r, α_r, β_r, and γ_r being linked together by the integral equation. It is precisely the labor involved in their determination that in practice often causes the problem to assume awesome proportions. By means of the results given herein, this labor is diminished to the irreducible minimum—the solving of a transcendental equation.

* Manuscript received by the PGIT, February 4, 1957. The work reported in this memo was performed in connection with Contract No. AF-18(600)-1505 with the Office of Sci. Res. of the Air Res. and Dev. Command.
† Polytechnic Inst. of Brooklyn, Brooklyn 1, N. Y.
[1] Subject to the usual modifications if the roots are not distinct.

Introduction

The integral equation

$$\phi(x) = \lambda \int_0^T K(x, y)\phi(y)\, dy, \qquad 0 \le x \le T \tag{1}$$

makes its appearance most naturally when one attempts to expand a random function $n(t, \omega)$ in an infinite series

$$n(t, \omega) = \sum_{k=1}^{\infty} a_k(\omega)\phi_k(t), \qquad k = 1, 2, \cdots, \cdots, \tag{2}$$
$$0 \le t \le T,$$

where the $\phi_k(t)$ are orthonormal over $0 \le t \le T$ and the $a_k(\omega)$ uncorrelated random variables.

Denote ensemble average by $E\{\ \}$ and let

$$m(t) = E\{n(t, \omega)\}$$

and

$$K(s, t) = E\{[n(s, \omega) - m(s)][n(t, \omega) - m(t)]\}.$$

Of course, $K(s, t)$ is the familiar covariance function of the process $n(t, \omega)$; if $K(s, t)$ is continous in the square $0 \le t \le T$, $0 \le s \le T$, the process is said to be second-order continuous in $0 \le t \le T$, and if in addition $K(s, t) = K(|s - t|)$, it is said to be second-order stationary.

From a theorem of Karhunen[2] any second-order random

[2] K. Karhunen, "Uber Lineare Methoden in der Wahrscheinlichkeitsrechnung," *Ann. Acad. Sci. Fennicae, Helsinki*, vol. 37; 1947.

function $n(t, \omega)$ which is second-order continuous in $0 \leq t \leq T$ may be expanded as follows:

$$n(t, \omega) = m(t) + \sum_{k=1}^{\infty} \frac{E_k(\omega)\phi_k(t)}{\sqrt{\lambda_k}}, \tag{3}$$

with convergence in the mean for every t in $[0, T]$. The quantities λ_k and $\phi_k(t)$ are determined from the integral equation

$$\phi_k(t) = \lambda_k \int_0^T K(s, t)\phi_k(s)\, ds, \qquad k = 1, 2, \cdots, \cdots. \tag{4}$$

The $\phi_k(t)$ is orthonormal over $0 \leq t \leq T$ and the $E_k(\omega)$ normalized uncorrelated random variables, *i.e.*,

$$E\{E_k(\omega)E_j(\omega)\} = \delta_{kj},$$

$$\int_0^T \phi_k(t)\phi_j(t)\, dt = \delta_{kj}, \qquad k, j = 1, 2, \cdots, \cdots$$

in which δ_{kj} is the Kronecker delta.

In particular if $n(t, \omega)$ is stationary, $K(s, t) = K(|s - t|)$ and (4) becomes

$$\phi_k(t) = \lambda_k \int_0^T K(|s - t|)\phi_k(s)\, ds, \tag{5}$$
$$k = 1, 2, \cdots, \cdots, \qquad 0 \leq t \leq T.$$

For Gaussian processes the $E_k(\omega)$ are actually independent normal variables and (3) converges in the ordinary sense with probability one. These properties of the expansion have been exploited to great advantage in many theoretical applications, for example,[3-6] the results therein fully justifying the importance of (4) and (5). The interested reader may find alternative treatments of (5) in Slepian,[5] DeSobrina,[7] and Muller.[8] As far as the author is aware, none of these treatments yield an explicit eigenvalue equation.

Notation and Assumptions

If what follows, we restrict ourselves to a special class of stationary processes. It follows directly from a theorem of Bochner that if $K(x, y)$ is the covariance function of a second-order stationary process with absolutely continuous spectral density, then

$$K(x, y) = K(|x - y|) = \int_{-\infty}^{+\infty} e^{i\omega(x-y)}G(\omega)\, d\omega, \tag{6}$$

where

$$G(\omega) = G(-\omega), \qquad G(\omega) \geq 0.$$

It is not difficult to show[9] that the eigenfunctions generated by such kernels by means of (5) are complete in $0 \leq x \leq T$, and that the eigenvalues are positive, their sole possible limiting point being infinity.

As is well-known if stationary noise with spectral density $G_1(\omega)$ is passed through a linear filter whose frequency response is $H(i\omega)$, then the output noise has spectral density $G_1(\omega) \mid H(i\omega) \mid^2$. For lumped constant systems

$$H(s) = \frac{h_0 + h_1 s + h_2 s^2 + , \cdots, + h_m s^m}{g_0 + g_1 s + g_2 s^2 + , \cdots, + g_n s^n},$$

$$s = i\omega, \quad h_j, \quad g_j \quad \text{real}.$$

For white input noise with mean square unity, $G_1(\omega) \equiv 1$ and the spectral density $G(\omega)$ of the output is

$$G(\omega) = H(i\omega)H(-i\omega) = \frac{N(-\omega^2)}{D(-\omega^2)} = \frac{N(s^2)}{D(s^2)}.$$

$N(s^2)$ and $D(s^2)$ being polynomials of degrees m and n in s^2, *viz.*,

$$N(s^2) = \sum_{k=0}^{m} a_{2k} s^{2k}, \tag{7}$$

$$D(s^2) = \sum_{k=0}^{n} b_{2k} s^{2k}, \qquad b_0 \neq 0.$$

The total output power equals

$$\int_{-\infty}^{+\infty} \frac{N(-\omega^2)}{D(-\omega^2)}\, d\omega.$$

In order that it be finite the degree of $D(s^2)$ must exceed that of $N(s^2)$ by at least two and $D(s^2)$ must have no roots on the imaginary s axis. Secondly, the requirement $G(\omega) \geq 0$, ω real, implies that any purely imaginary zero of $N(s^2)$ must be of even multiplicity.

The inverse of $(s^2 - \alpha^2)^{-k}$, for k a positive integer and the real part of $\alpha > 0$ is

$$\frac{(-1)^k}{(k-1)!} e^{-\alpha|x|} \sum_{j=1}^{k} \frac{(k + j + 2)!\, |x|^{k-j}}{(j-1)!\,(k-j)!\,(2\alpha)^{k+j-1}}.$$

By expanding $N(s^2)/D(s^2)$ in partial fractions it is seen that $K(|x - y|)$ must be the sum of products of polynomials in $|x - y|$ with exponentials of $|x - y|$.

Denote the roots of $D(s^2)$ by $\pm \mu_1, \pm \mu_2, \cdots, \pm \mu_n$ and assume them so ordered that $0 < \text{Re}\, \mu_1 \leq \text{Re}\, \mu_2 \leq, \cdots, \leq \text{Re}\, \mu_n$. Define the polynomials $D^+(s)$ and $D^-(s)$ by $D(s^2) = D^+(s)\, D^-(s)$, $D^+(-s) = D^-(s)$, the roots of $D^+(s)$ being $-\mu_j$ and those of $D^-(s)$, $+\mu_j$, $j = 1, 2, \cdots, n$. Symbolically,

$$D^+(s) = \sum_{k=0}^{n} d_k s^k, \qquad d_k \quad \text{real}.$$

[3] R. C. Davis, "The detectability of random signals in the presence of noise," IRE Trans., vol. IT-3, pp. 52–62; March, 1954.
[4] U. Grenander, "Stochastic processes and statistical inference," *Ark F. Mat.*, vol. 1, pp. 195–277; October, 1950.
[5] D. Slepian, "Estimation of signal parameter in the presence of noise," IRE Trans., vol. IT-3, pp. 68–89; March, 1954.
[6] D. C. Youla, "The use of the method of maximum likelihood in estimating continuous modulated intelligence which has been corrupted by noise," IRE Trans., vol. IT-3, pp. 90–105; March, 1954.
[7] R. DeSobrino, "Optimum signal detection with incompletely specified signal noise," Ph. D. dissertation, Columbia Univ., New York, N. Y.; April, 1953.
[8] F. A. Muller, "Communication in the Presence of Additive Gaussian Noise," Tech. Rep. No. 244, M.I.T. Res. Lab. of Electronics; May, 1953.
[9] D. C. Youla, "A Finite-Time Homogeneous Weiner Hopf Integral Equation," Tech. Rep. No. 367, Microwave Res. Inst., Polytechnic Inst. of Brooklyn, N.Y.; October, 1955.

The following properties of $K(x)$ may be derived without too much effort from its representation in (6) and the fact that its transform is a rational function of ω^2 possessing the properties enumerated above.

1) The highest order derivative possessed by $K(x)$ at the origin is $2n - 2m - 2$.

2) $$D^+\left(\frac{d}{dx}\right)K(x) = 0, \qquad x > 0.$$

3) $$D^-\left(\frac{d}{dx}\right)K(x) = 0, \qquad x < 0.$$

4) $$D\left(\frac{d^2}{dx^2}\right)K(x) = 0, \qquad x \neq 0.$$

5) $K(x)$ possesses continuous derivatives of all order for $x \neq 0$.

The Determination of the Eigenvalues and Eigenfunctions Generated by a Kernel $K(|x-y|)$ whose Transform is a Rational Function of s^2

We now direct our attention exclusively to the integral equation

$$\phi(x) = \lambda \int_0^T K(x - y)\phi(y)\, dy, \tag{8}$$
$$0 \leq x \leq T,$$

where

$$K(x) \doteqdot \frac{N(s^2)}{D(s^2)}, \qquad -\text{Re}\,\mu_1 < \text{Re}\, s < \text{Re}\,\mu_1$$

and $N(s^2)$ and $D(s^2)$ are as described in (6) and (7). The notation is that employed by van der Pol.[10] From the breakup

$$\phi(x) = \lambda \int_0^x K(x - y)\phi(y)\, dy + \lambda \int_x^T K(x - y)\phi(y)\, dy \tag{9}$$

of (8) and the continuity properties of $K(x)$, it follows that $\phi(x)$ possesses continuous derivatives of all orders for $0 < x < T$. Again by the continuity properties of $K(x)$ and its derivatives, it follows that the function

$$g(x) \equiv \int_0^T K(x - y)\phi(y)\, dy, \qquad -\infty < x < \infty,$$

where $\phi(x)$ is a solution of (8), possesses continuous derivatives of all orders everywhere except perhaps at $x = 0$ and $x = T$; for $x < 0$ or $x > T$ these derivatives are given by

$$\frac{d^r g}{dx^r} = \int_0^T \frac{\partial^r K(x - y)}{\partial x^r}\phi(y)\, dy, \qquad r = 1, 2, \cdots, \cdots.$$

Lastly,

$$\frac{d^r g}{dx^r} = 0[e^{-(\text{Re}\,\mu_1 + \delta)|x|}] \quad \text{as} \quad |x| \to \infty,$$
$$(r = 0, 1, \cdots, \cdots), \qquad \delta > 0.$$

Define $\Phi(x)$ by

$$\Phi(x) = \phi(x), \qquad 0 \leq x \leq T,$$
$$= 0 \quad \text{otherwise}$$

and let

$$v(x) \equiv \Phi(x) - \lambda g(x),$$
$$= \Phi(x) - \lambda \int_0^T K(x - y)\Phi(y)\, dy, \tag{10}$$
$$-\infty < x < \infty.$$

From the previous discussion, the properties 2) and 3) of $K(x)$ and the fact that $\phi(x)$ is a solution of (8), it follows that

1) $$D^+\left(\frac{d}{dx}\right)v(x) = 0, \qquad x > T,$$

2) $$D^-\left(\frac{d}{dx}\right)v(x) = 0, \qquad x < T,$$

3) $v(x) = 0, \quad 0 \leq x \leq T$ and

4) The left and right hand limits of $v(x)$ as $x \to 0$, T, exist and are finite.

As an immediate consequence of these properties, it follows that $V(s)$, the bilateral Laplace transform of $v(x)$ is given by

$$V(s) = \frac{P(s)}{D^-(s)} - e^{-sT}\frac{Q(s)}{D^+(s)},$$
$$-\text{Re}\,\mu_1 < \text{Re}\, s < \text{Re}\,\mu_1, \tag{11}$$

$P(s)$ and $Q(s)$ being polynomials of degree $n - 1$ at most. The proof is obvious once it is realized that $v(x)$ is a sum of exponentials in both the regions $x > T$ and $x < 0$.

Denote the bilateral Laplace transform of $\Phi(x)$ by $\overline{\Phi}(s)$. Transforming both sides of (10) with respect to x (a common strip of convergence exists) yields

$$V(s) = \overline{\Phi}(s)\left[1 - \lambda\frac{N(s^2)}{D(s^2)}\right]$$
$$-\text{Re}\,\mu_1 < \text{Re}\, s < \text{Re}\,\mu_1.$$

Using (11),

$$\overline{\Phi}(s) = \frac{D^+(s)P(s) - e^{-sT}D^-(s)Q(s)}{D(s^2) - \lambda N(s^2)}, \tag{12}$$
$$-\text{Re}\,\mu_1 < \text{Re}\, s < \text{Re}\,\mu_1.$$

However, since $\overline{\Phi}(s)$ is an integral function of s,

$$\left(\frac{d\overline{\Phi}}{ds} = -\int_0^T x\phi(x)e^{-sx}\, dx\right),$$

[10] B. van der Pol and H. Bremmer, "Operational Calculus Based on the Two Sided Laplace Transform," Cambridge University Press, Cambridge, Eng ; 1955.

(12) is valid in the entire strip $-\infty < \text{Re } s < \infty$ and numerator and denominator must possess the same order zeros.

In (12) we have an explicit representation for the transform of an eigenfunction. It remains to establish the connection between $P(s)$ and $Q(s)$.

Let the roots of $D(s^2, \lambda) \equiv D(s^2) - \lambda N(s^2)$ be $\pm\,\omega_1(\lambda)$, $\pm\,\omega_2(\lambda)$, $\pm, \cdots, \pm\,\omega_n(\lambda)$ arranged so that $0 \leq \text{Re } \omega_1 \leq \text{Re } \omega_2 \leq, \cdots, \leq \text{Re } \omega_n$. In what follows we shall, in order to minimize the algebra, assume that λ is an eigenvalue for which the $\omega_k(\lambda)$ are distinct. In any case there exist at most a finite number of eigenvalues which violate this assumption and they are easily determined in any given problem.

By equating the numerator of (12) to zero for $s = \pm\,\omega_1(\lambda), \pm\,\omega_2(\lambda), \pm, \cdots, \pm\,\omega_n(\lambda)$ we get the n pairs of equations

$$\begin{aligned} D^+(\omega_r)P(\omega_r) &= e^{-\omega_r T}D^-(\omega_r)Q(\omega_r), \\ D^-(\omega_r)P(-\omega_r) &= e^{\omega_r T}D^+(\omega_r)Q(-\omega_r), \\ r &= 1, 2, \cdots, n. \end{aligned} \tag{13}$$

Multiplication of corresponding sides of (13) yields [any factor common to both $D(s^2)$ and $N(s^2)$ is assumed to have been cancelled],

$$\begin{aligned} P(\omega_r)P(-\omega_r) &= Q(\omega_r)Q(-\omega_r), \\ r &= 1, 2, \cdots, n. \end{aligned} \tag{14}$$

The polynomial $L(s^2) \equiv P(s)\,P(-s) - Q(s)\,Q(-s)$ is of degree $n - 1$ in s^2 and vanishes for n distinct values of s^2, namely $\omega_1^2, \omega_2^2, \cdots, \omega_n^2$. Hence

$$P(s)P(-s) \equiv Q(s)Q(-s). \tag{15}$$

Let $P(s) = p_{n-1}\,(s - \alpha_1)(s - \alpha_2) \cdots (s - \alpha_{n-1})$. Then any solution $Q(s)$ of (15) is of the form ($j_1, j_2, \cdots, j_{n-1}$ is any permutation of $1, 2, \cdots, n - 1$)

$$\begin{gathered} Q(s) = \pm p_{n-1} \prod_{k=1}^{r} (s - \alpha_{j_k}) \prod_{l=1}^{n-r-1} (s + \alpha_{j_{r+l}}), \\ 0 \leq r \leq n - 1, \end{gathered} \tag{16}$$

and are 2^n in number. But from (13) it is seen that the coefficients $q_k(k = 0, 1, \cdots, n - 1)$ of $Q(s)$ must depend linearly on the coefficients $p_k(k = 0, 1, \cdots, n - 1)$ of $P(s)$. This allows us to rule out all but four solutions of (16), namely, $P(s) = \pm\,Q(s)$ and $P(s) = \pm\,Q(-s)$.

The following example should make this clear. Suppose

$$P(s) = p_0 + p_1 s + p_2 s^2$$

and

$$Q(s) = q_0 + q_1 s + q_2 s^2.$$

Then

$$P(s)P(-s) = p_0^2 + (2p_0p_1 - p_1^2)s + p_2^2 s^4$$

and

$$Q(s)Q(-s) = q_0^2 + (2q_0q_1 - q_1^2)s + q_2^2 s^4,$$

so that by (15)

$$p_0^2 = q_0^2, \qquad p_2^2 = q_2^2,$$
$$2p_0p_2 - p_1^2 = 2q_0q_2 - q_1^2\,.$$

The solutions are

$p_0 = -\,q_0$	$p_0 = -q_0$	$p_0 = -q_0$	$p_0 = q_0$
$p_2 = -q_2$	$p_2 = q_2$	$p_2 = -q_2$	$p_2 = q_2$
$p_1 = \pm\sqrt{q_1^2 - 4q_0q_2}$	$p_1 = \pm\sqrt{q_1^2 - 4q_0q_2}$	$p_1 = \pm q_1$	$p_1 = \pm q_1\cdot$

To prove that $P(s) = \pm\,Q(s)$ and $P(s) = \pm\,Q(-s)$ are the only linear solutions of (16), note first that they correspond to $r = n - 1$ and $r = 0$, respectively; the others correspond to $0 < r < n - 1$. The coefficient of s^{n-2} in $Q(s)$ is, for any r, given by

$$\begin{aligned} q_{n-2} = \mp p_{n-1}(\alpha_{j_1} + \alpha_{j_2} +, \cdots, + \alpha_{j_r} - \alpha_{j_{r+1}} \\ - \alpha_{j_{r+2}} -, \cdots, - \alpha_{j_{n-1}}). \end{aligned}$$

Clearly if $r \neq 0$ or $n - 1$, q_{n-2} is not a symmetric function of the roots $\alpha_1, \alpha_2, \cdots, \alpha_{n-1}$ of $P(s)$ and this in turn implies, *a fortiori*, that it cannot be a linear combination of the p_k.

The two pairs of solutions $P(s) = \pm\,Q(-s)$, $P(s) = \pm\,Q(s)$, coincide when $n = 1$. A glance at (13) reveals that the former pair is consistent with both the upper and lower sets of equations and that for $n > 1$ the latter pair is not. Thus the only possible relations between the polynomials P and Q are $P(s) = \pm\,Q(-s)$. Substituting for $Q(s)$ in the top set in (13) we get,

$$\sum_{k=0}^{n-1} [1 \mp (-1)^k x_r]\omega_r^k p_k = 0, \tag{17}$$

where

$$x_r \equiv e^{-\omega_r T}\left[\frac{D^-(\omega_r)}{D^+(\omega_r)}\right], \qquad (r = 1, 2, \cdots, n).$$

In order that a nontrivial solution for the p_k exist, the determinant of the system must equal zero. Hence,

$$\Delta = \begin{vmatrix} (1 \mp x_1), & (1 \pm x_1)\omega_1 & \cdots, & [1 \mp (-1)^{n-1}x_1]\omega_1^{n-1} \\ (1 \mp x_2), & (1 \pm x_2)\omega_2 & \cdots, & \cdot \\ \vdots & \vdots & \vdots & \vdots \\ (1 \mp x_n), & (1 \pm x_n)\omega_n & \cdots, & [1 \mp (-1)^{n-1}x_n]\omega_n^{n-1} \end{vmatrix} = 0. \tag{18}$$

These two transcendental equations serve to determine the eigenvalues which are contained implicitly in the x's and ω's.

When $n = 1$, (18) reduces to $x_1 = \pm 1$, or

$$e^{\omega_1 T} = \pm\frac{D^-(\omega_1)}{D^+(\omega_1)}. \tag{19}$$

A straightforward application of the Maximum Modulus Theorem reveals that for any $n \geq 0$ all solutions of

For $n > 1$ (18) may be written in a more manageable form as (we choose $n = 3$ to avoid cumbersome determinants)

$$\Delta = \begin{vmatrix} 1 & \omega_1 \coth \theta_1 & \omega_1^2 \tanh \theta_1 \\ 1 & \omega_2 \coth \theta_2 & \omega_2^2 \tanh \theta_2 \\ 1 & \omega_3 \coth \theta_3 & \omega_3^2 \tanh \theta_3 \end{vmatrix} = 0, \tag{24}$$

where

$$\tanh \theta_r = \frac{(d_0 + d_2\omega_r^2 + \cdots) \tanh (\omega_r T/2) + (d_1\omega_r + d_3\omega_r^3 + \cdots)}{(d_0 + d_2\omega_r^2 + \cdots) + (d_1\omega_r + d_3\omega_r^3 + \cdots) \tanh (\omega_r T/2)}, \qquad r = 1, 2, \cdots, n. \tag{25}$$

$$e^{sT} = \pm\frac{D^+(s)}{D^-(s)} \tag{20}$$

lie on the purely imaginary s axis. Denote them by $\pm i\beta_k(k = 0, 1, \cdots, \cdots)$. After a little algebra we find that the roots of (20) may be determined from the two transcendentals

$$\begin{aligned} \tan \frac{\beta T}{2} &= -\frac{d_1\beta - d_3\beta^3 + d_5\beta^5 - , \cdots}{d_0 - d_2\beta^2 + d_4\beta^4 - , \cdots}, \\ \operatorname{ctn} \frac{\beta T}{2} &= \frac{d_1\beta - d_3\beta^3 + d_5\beta^5 - , \cdots}{d_0 - d_2\beta^2 + d_4\beta^4 - , \cdots}. \end{aligned} \tag{21}$$

Now when

$$\begin{aligned} n = 1, \qquad N(s^2) &= a_0 > 0, \\ D(s^2) &= b_0 - cs^2, \qquad b_0, c > 0, \\ \omega_1(\lambda) &= i\sqrt{\frac{a_0}{c}\left(\lambda - \frac{b_0}{a_0}\right)}, \\ D^+(s) &= \sqrt{b_0} + \sqrt{c}\, s \end{aligned} \tag{21}$$

and reduce to

$$\tan \frac{\beta T}{2} = -\sqrt{\frac{c}{b_0}},$$

and

$$\operatorname{ctn} \frac{\beta T}{2} = \sqrt{\frac{c}{b_0}}\, \beta. \tag{22}$$

The λ_k are calculated from

$$\lambda_k = \frac{b_0 + c\beta_k^2}{a_0}, \qquad (k = 0, 1, \cdots, \cdots). \tag{23}$$

The other equation is obtained by interchanging $\tanh \theta_r$ and $\coth \theta_r$ in (24).

Once the eigenvalues λ_r have been determined, the p_k are found from (17) and through them $\phi_r(x)$. Explicitly,

$$\phi_r(x) \doteqdot \frac{P(s, \lambda_r)D^+(s)}{D(s^2) - \lambda_r N(s^2)}, \tag{26}$$

$$\operatorname{Re} s > \omega_r(\lambda_r),$$

$$0 \leq x \leq T.$$

The ϕ_r defined by (26) are orthogonal but not orthonormal over $0 \leq x \leq T$. An expression for the normalizing constant can probably be derived but its undoubted complexity makes it useless for practical work.

The Semi-Infinite Range Singular Integral Equation

The solution of the homogeneous integral equation

$$\phi(x) = \lambda \int_0^\infty K(x - y)\phi(y)\, dy, \qquad 0 \leq x < \infty \tag{27}$$

has been given by Wiener and Hopf for a wide class of kernels and can be found reproduced in Titchmarsh's book.[11] However, when the transform of $K(x)$ is a rational function of s^2 our elementary technique gives the solution immediately. Omitting the details and adhering to the same notation as before, we find that

[11] E. C. Titchmarsh, "Introduction to the Theory of Fourier Integrals," Clarendon Press, Oxford, Eng., 2nd ed.; 1948.

$$\phi''(0^+, \lambda) = \frac{b_{2n}(d_n p_{n-3} + d_{n-1}p_{n-2} + d_{n-2}p_{n-1}) - (b_{2n-2} - \lambda A_{2n-2})\, d_n p_{n-1}}{b_{2n}^2}$$

$$\bar{\Phi}(s) = \frac{P(s)D^-(s)}{D(s^2) - \lambda N(s^2)}, \tag{28}$$

$$\operatorname{Re} s > \operatorname{Re} \omega_n(\lambda),$$

the polynomial $P(s)$ being arbitrary and of degree $n - 1$. The eigenvalues are determined by the requirement that the integral in (29) converge. This leads to

$$\operatorname{Re} \omega_n(\lambda) < \operatorname{Re} \mu_1. \tag{29}$$

All λ satisfying (29) are eigenvalues.

The coefficients p_k of $P(s)$ serve to fix the initial values $\phi(0^+, \lambda)$, $\phi'(0^+, \lambda)$, $\cdots$, $\phi^{n-1}(0^+, \lambda)$. Since the number of linearly independent polynomials $P(s)$ is n, each eigenvalue has n-fold degeneracy. To give a formula for $\phi^r(0^+, \lambda)$, $(r = 0, 1, \cdots, n-1)$ in terms of the p_k, a_{2k}, and b_{2k}, it is first necessary to introduce some preliminary notation. Let

$$P(s)D^+(s) \equiv \sum_{k=0}^{2n-1} A_k s^k$$

and

$$D(s^2) - \lambda N(s^2) \equiv \sum_{k=0}^{n} B_{2k} s^{2k},$$

where

$$A_k = \sum_{j=0}^{k} p_{k-j}\, d_j$$

and

$$B_{2k} = b_{2k} - \lambda A_{2k}.$$

It is understood that $A_{2k} = 0$ for $k > m$ and any p or d whose subscript exceeds $n - 1$ and n, respectively, is to be taken equal to zero. Then (again omitting details),

$$\phi^r(0^+, \lambda) = (-1)^r \frac{\Omega^{r+3}}{b_{2n}^{r+3}}, \qquad r = 0, 1, 2, \cdots, \cdots, \tag{30}$$

where

$$\Omega^3 = \begin{vmatrix} 0 & B_{2n} & 0 \\ 0 & 0 & B_{2n} \\ A_{2n-1} & B_{2n-2} & 0 \end{vmatrix},$$

$$\Omega^4 = \begin{vmatrix} 0 & B_{2n} & 0 & 0 \\ 0 & 0 & B_{2n} & 0 \\ A_{2n-1} & B_{2n-2} & 0 & B_{2n} \end{vmatrix} \text{ etc.} \tag{31}$$

For example,

$$\phi(0^+, \lambda) = \frac{d_n p_{n-1}}{b_{2n}},$$

$$\phi'(0^+, \lambda) = \frac{d_n p_{n-2} + d_{n-1}p_{n-1}}{b_{2n}}, \tag{32}$$

etc.

Illustrative Examples

Consider the Picard kernel

$$K(x) = \delta^2 e^{-k|x|} \doteqdot \frac{2k\delta^2}{k^2 - s^2}, \qquad k, \delta^2 > 0,$$

$$-k < \operatorname{Re} s < k.$$

Here $a_0 = 2k\delta^2$, $b_0 = k^2$, and $c = 1$. From (22) and (23) we get

$$\tan \frac{\beta T}{2} = -\beta/k,$$

$$\operatorname{ctn} \frac{\beta T}{2} = \beta/k,$$

$$\lambda_r = \frac{k^2 + \beta_r^2}{2k\delta^2}$$

and

$$\phi_r(x) = \frac{k}{\beta_r} \sin \beta_r x + \cos \beta_r x.$$

The $\phi_r(x)$ is not normalized.

A more complicated specimen is the kernel

$$K(x) \doteqdot \frac{1}{s^4 + 1}, \qquad -\frac{1}{\sqrt{2}} < \operatorname{Re} s < \frac{1}{\sqrt{2}}.$$

$$\therefore D^+(s) = s^2 + \sqrt{2}s + 1,$$

$$D^-(s) = s^2 - \sqrt{2}s + 1,$$

and

$$D(s^2) - \lambda N(s^2) = s^4 - (\lambda - 1).$$

Let

$$\omega_2(\lambda) = (\lambda - 1)^{\frac{1}{4}} \equiv \epsilon$$

and

$$\omega_1(\lambda) = i\epsilon$$

where Im. $\epsilon \leq 0 \leq$ Re. ϵ. From (24), the eigenvalue equations are found to be

$$\omega_2 \coth \theta_2 = \omega_1 \coth \theta_1,$$

and

$$\omega_2 \tanh \theta_2 = \omega_1 \tanh \theta_1$$

in which

$$\tanh \theta_1 = \frac{(1 + \omega_1^2) \tanh (\omega_1 T/2) + \sqrt{2}\omega_1}{(1 + \omega_1^2) + \sqrt{2}\omega_1 \tanh (\omega_1 T/2)},$$

$$\tanh \theta_2 = \frac{(1 + \omega_2^2) \tanh(\omega_2 T/2) + \sqrt{2}\omega_2}{(1 + \omega_2^2) + \sqrt{2}\omega_2 \tanh (\omega_2 T/2)}.$$

fter replacing ω_1 and ω_2 by their values in terms of ϵ they mplify to

$$\frac{\sqrt{2}\epsilon + (1+\epsilon^2)\tanh(\epsilon T/2)}{(1+\epsilon^2) + \sqrt{2}\epsilon\tanh(\epsilon T/2)}$$

$$= \pm\frac{\sqrt{2}\epsilon + (1-\epsilon^2)\tan(\epsilon T/2)}{(1-\epsilon^2) - \sqrt{2}\epsilon\tan(\epsilon T/2)}. \qquad (33)$$

Denote their solutions by $\epsilon_1, \epsilon_2, \cdots, \cdots$, the odd ibscripts corresponding to the plus sign. Then, $\lambda_r = \epsilon_r^4 + 1$, $(r = 1, 2, \cdots, \cdots)$; from (17),

$$p_{1r} = -\frac{p_0}{\epsilon_r}\tanh\theta_1(\epsilon_r), \qquad r \text{ odd}$$

$$= -\frac{p_0}{\epsilon_r}\coth\theta_1(\epsilon_r), \qquad r \text{ even}$$

nd

$$(x) \doteqdot \frac{(s^2 + \sqrt{2}s + 1)\left[1 - \frac{s}{\epsilon_r}\tanh\theta_1(\epsilon_r)\right]}{s^4 - \epsilon_r^4}, \qquad r \text{ odd},$$

$$\doteqdot \frac{(s^2 + \sqrt{2}s + 1)\left[1 - \frac{s}{\epsilon_r}\coth\theta_1(\epsilon_r)\right]}{s^4 - \epsilon_r^4}, \qquad r \text{ even},$$

$$\operatorname{Re} s > \epsilon_r,$$

$$0 \le x \le T.$$

If this same kernel is used over the semi-infinite range the eigenfunctions are most simple, *viz.*,

$$\phi(x, \lambda) \doteqdot \frac{P(s)(s^2 + \sqrt{2}s + 1)}{s^4 + (1-\lambda)}, \qquad 0 \le x \le \infty.$$

To find the eigenvalues set $\lambda - 1 = r^4 e^{i\rho}$. Then, $\omega_2(\lambda) = r\, e^{i\rho/4}(0 \le \rho < 2\pi)$ and from (29) $r\cos(\rho/4) < 1/\sqrt{2}$; a point (r, ρ) in this region determines the eigenvalue $\lambda(r, \rho) = 1 + r^4 e^{i\rho}$. Hence the strip of covergence of the above transform is Re $s > r\cos(\rho/4)$. Note that now the eigenvalues are no longer discrete but form a continuum. This is not surprising since K is not square-integrable over $0 \le x, y < \infty$ and, consequently, the theory of bounded linear symmetric operators is inapplicable.

ACKNOWLEDGMENT

The author wishes to take this opportunity to thank Dr. D. Slepian for his very careful reading of the original draft of this paper and his many helpful comments.

The Correlation Function of Smoothly Limited Gaussian Noise*

R. F. BAUM†

Summary—**The correlation function of "smoothly" limited Gaussian noise is calculated and compared with the correlation function of "extremely" clipped Gaussian noise. The limiting function is assumed to have the shape of the error integral curve. The output spectrum is calculated for the case of noise passed through an RC filter.**

I. INTRODUCTION

THE effect of clipping on Gaussian noise has been investigated in detail.[1-3] Van Vleck has analyzed the case of an amplitude limiter with an output vs input characteristic $y(x)$ described by

$$\begin{aligned} y(x) &= x && \text{for } -x_0 < x < +x_0 \\ &= x_0 && \text{for } x > x_0 \\ &= -x_0 && \text{for } x < -x_0 \end{aligned}$$

and for varing ratios of rms noise to clipping threshold x_0. As this ratio approaches infinity the case of extreme clipping is reached.

Instead of clipping suddenly, when x exceeds the threshold x_0, smooth or gradual limiting may be preferred. The limiter characteristic then will have the aspect of Fig. 1(a). Van Vleck has calculated the autocorrelation function of the output noise for a limiter of the shape

$$y(x) = c_1 \tanh^{-1}(c_2 x).$$

Unfortunately the result cannot be expressed in closed form. Another example is discussed in Laning and Battin.[3]

The present paper assumes a limiter of the form of an error-integral curve. This curve has considerable practical importance because:

* Manuscript received by the PGIT, March 13, 1957.
† The W. L. Maxson Corporation, New York, N.Y.
1 J. H. Van Vleck, "The spectrum of clipped noise," RRL Rep. 51; July 21, 1943.
2 S. L. Lawson and G. E. Uhlenbeck, "Threshold Signals," McGraw-Hill Book Co., Inc., New York, N.Y., p. 58; 1950.
3 J. H. Laning and R. H. Battin, "Random Processes in Automatic Control," McGraw-Hill Book Co., Inc., New York, N. Y., p. 163; 1956.

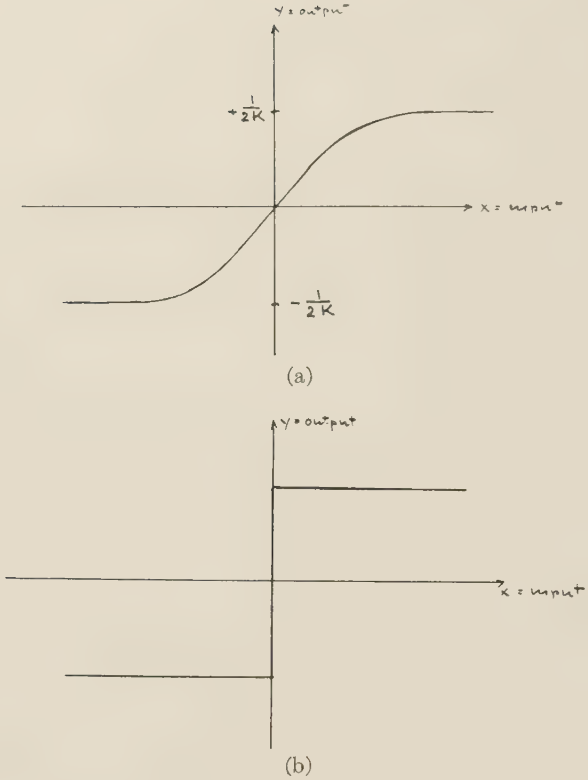


Fig. 1—Limiter curve. (a) $\sigma_{0s} \neq 0$; (b) $\sigma_{0s} = 0$.

(a) The probability density distribution of the output noise can be made constant over a prescribed range of output voltage. The spikes of the distribution curve, at the threshold level of which accompany sudden clipping, disappear. The output voltage spends the same average time at any voltage level within the range.

(b) Therefore, if the output noise is used to fm modulate an rf carrier, the output power can be evenly distributed over a sharply defined frequency band.

The output correlation function appears mathematically in closed form, which facilitates the calculation of the output spectrum for a prescribed input spectrum. As an example, the output spectrum for Gaussian noise passed through an RC filter is found and compared with the spectrum of "extremely" clipped noise.

II. Correlation Function and Spectrum After Smooth Limiting

The smooth clipping function is defined by

$$y(x) = \frac{1}{K} \frac{1}{\sqrt{2\pi\sigma_{0s}}} \int_0^x \epsilon^{-z^2/2\sigma_{0s}} \, dz \qquad (1)$$

where x is the input voltage and K is a constant. Since this function is proportional to the integral of a normalized Gaussian probability curve of rms value σ_{0s}, this parameter will be called rms value of the limiter curve, as distinguished from the rms value σ_0 of the input noise. The parameter

$$\alpha = \frac{\sigma_{0s}}{\sigma_0} \qquad (2)$$

then defines the ratio of the two rms values.

The shape of the limiting curve is indicated in Fig. 1(a). The curve limits the noise to values between plus and minus $1/2K$. The special case of extreme clipping shown in Fig. 1(b) is obtained from the result as σ_{0s} approaches zero.

The particular choice of this function is explained by the fact that for $\alpha = 1$ the input probability distribution P_{in} [Fig. 2(a)] is converted to the output probability distribution P_{out} [Fig. 2(b)], which is constant between the values $\pm\ 1/2K$.

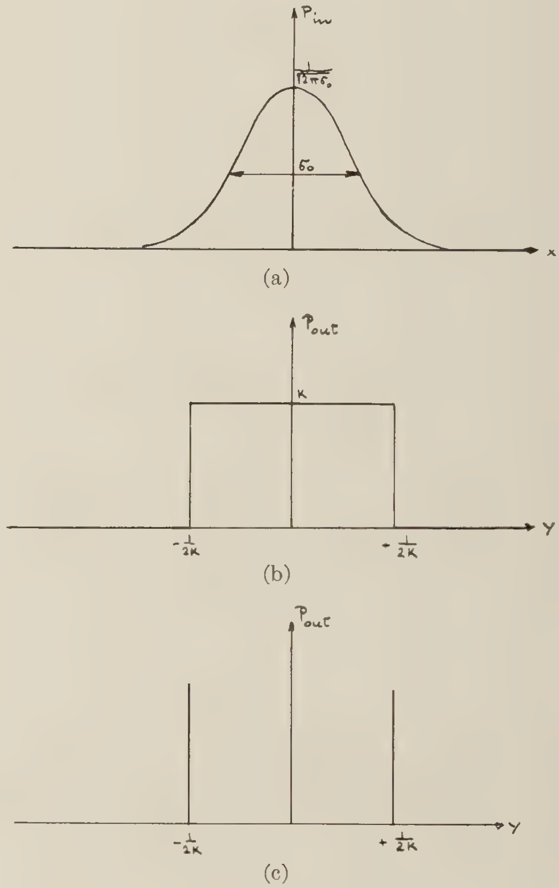


Fig. 2—Probability distributions. (a) Gaussian (input), (b) smoothly limited ($\alpha = 1$), (c) extremely clipped ($\alpha = 0$).

The output probability distribution can easily be established by application of the well-known formula

$$P_{\text{out}} = \frac{P_{\text{in}}}{dy/dx} \qquad (3)$$

by using the derivative of y from (1) and with

$$P_{\text{in}} = \frac{1}{\sqrt{2\pi\sigma_0}} \epsilon^{-x^2/2\sigma_0}. \qquad (4)$$

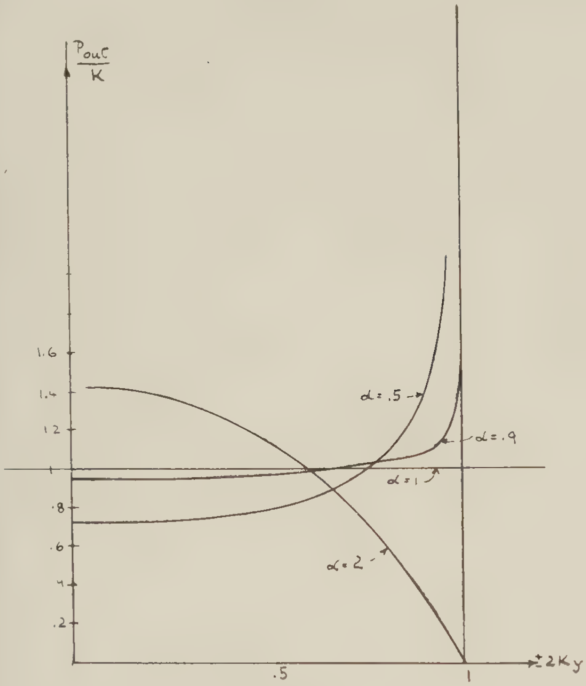


Fig. 3—Probability distribution for varying input levels.

This results in

$$P_{\text{out}}(x) = K\sqrt{\alpha}\,\epsilon^{-(x^2/2\sigma_0)(\alpha-1)}. \tag{5}$$

Since x and y are related by (1), this also gives P_{out} as a function of y. In Fig. 3, P_{out}/K is plotted as a function of $2Ky$ with different values of α as parameter.

It is interesting to note that at $2Ky = 1$, P_{out} remains zero for $\alpha < 1$ but develops an infinite peak for $\alpha > 1$. As $\alpha \to 0$ (extreme clipping) only these peaks remain [Fig. 2(c)] since the output noise can have no other values but $y = \pm 1/2K$. Furthermore, as (5) shows, for $y = x = 0$

$$\frac{P(0)}{K} = \sqrt{\alpha}. \tag{6}$$

If the normalized input correlation function ρ_τ is given, the normalized output correlation function P_τ can be found. The details of the calculation are given in Appendix A. The result is

$$\mathrm{P}_\tau = \frac{\sin^{-1}\dfrac{\rho_\tau}{1+\alpha}}{\sin^{-1}\dfrac{1}{1+\alpha}}. \tag{7}$$

In particular for $\alpha \to 0$ (extreme clipping) one obtains

$$\mathrm{P}_\tau = \frac{2}{\pi}\sin^{-1}\rho_\tau \tag{8}$$

which checks with the value obtained in Van Vleck.[1] Also, for $\alpha = 1$ (constant output probability distribution):

$$\mathrm{P}_\tau = \frac{6}{\pi}\sin^{-1}\left(\frac{\rho_\tau}{2}\right). \tag{9}$$

For $\alpha \to \infty$ (amplification) we obtain the obvious result $\mathrm{P}_\tau = \rho_\tau$.

The output spectrum is the cos transform of the autocorrelation function:

$$S(\omega) = \frac{4}{\sin^{-1}\left(\dfrac{1}{1+\alpha}\right)}\int_0^\infty \sin^{-1}\left(\frac{\rho_\tau}{1+\alpha}\right)\cos\omega\tau\,d\tau. \tag{10}$$

An evaluation of $S(w)$ for the case $\alpha = 1$ is given in Appendix B for Gaussian noise passed through RC filter of bandwidth B, for which

$$\rho_\tau = \epsilon^{\;B\tau}. \tag{11}$$

One may expect from intuition that the harmonic content of smoothly limited noise is far less than that of extremely clipped noise. In general the output spectrum can be thought of as resulting from the superposition of an infinite number of spectra extending individually from zero to B, 3B, 5B, $\cdots$ etc., as shown in Fig. 4.

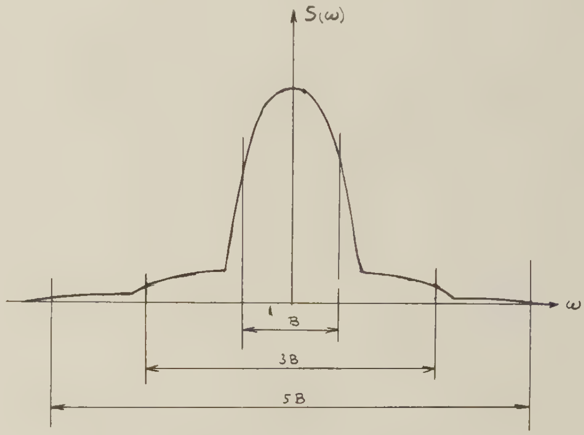


Fig. 4—Output spectrum after limiting.

Table I compares the amplitude of the first three spectra for smoothly limited and extremely clipped Gaussian noise. The amplitude of the basic output spectrum of bandwidth B is $3/\pi$ and $2/\pi$, respectively, as shown in the first column. The last column gives the sum of the first three spectra. This equals, for $\alpha = 1$, almost one. Therefore the power outside the band B is quite small.

Finally it is pointed out that (7) can also be used to calculate the input correlation function ρ_τ, if an output

TABLE I

	B	3B	5B	Sum
Smoothly Limited $\alpha = 1$	0.9549	0.0398	0.0045	0.9993
Extremely Clipped $\alpha = 0$	0.6366	0.1571	0.0478	0.8415

correlation function P_τ is prescribed. For instance, it could be asked if it is possible to completely eliminate the spread of the spectrum by prescribing an output correlation function of the form of (11). The answer is negative, since the calculation leads to negative terms in the expression for the necessary input spectrum.

Appendix

A. Evaluation of the Normalized Output Autocorrelation Function

If the input autocorrelation function φ_τ is given, the output autocorrelation function of limited Gaussian noise is given by the integral[4]

$$\psi_\tau = \frac{1}{4\pi^2}\int_C F(ju)\int_C F(jv)\epsilon^{-\sigma_0/2(u^2+v^2)-uv\varphi_\tau}\,du\,dv \tag{12}$$

where $F(ju)$ is determined by the limiting function $y(x)$:

$$y(x) = \frac{1}{2\pi}\int_C F(ju)\epsilon^{ixu}\,du. \tag{13}$$

If we use the imaginary axis for the contour C, $F(ju)$ becomes the Fourier transform of $y(x)$.

In the following analysis we shall substitute for the limiting function $y(x)$, the limiting function $y^*(x)$:

$$y^*(x) = y(x) + \frac{1}{2K},$$

the additional constant is obviously chosen so as to make $y^* \to 0$ as $x \to -\infty$. The Fourier transform of $y^*(x)$ is[5]

$$F(ju) = \frac{1}{juK}\epsilon^{-(\sigma_{0s}/2)u^2}. \tag{14}$$

The use of y^* instead of y will merely add a dc voltage to the time varying output voltage, which otherwise maintains its shape. Therefore, the output correlation function will contain a constant term, due to the dc voltage, and a time dependent term, due to noise, which is the same for $y^*(x)$ and $y(x)$. As the calculation shows, the constant term appears in the form of an integration constant (21), which therefore has to be disregarded. Because of the symmetrical limiter curve the input noise by itself cannot account for a dc component in the output.

By substitution of (14) into (12) the integral becomes

$$\psi_\tau = \frac{1}{4\pi^2K^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{1}{uv}\epsilon^{-[(\sigma_0+\sigma_{0s})/2](u^2+v^2)-uv\varphi_\tau}\,du\,dv. \tag{15}$$

We may eliminate the factor $1/uv$ by first calculating the derivative of ψ_τ with respect to φ_τ and simplify further by introducing the new variables:

$$u' = \sqrt{\frac{\sigma_0+\sigma_{0s}}{2}}\,u$$

$$v' = \sqrt{\frac{\sigma_0+\sigma_{0s}}{2}}\,v.$$

This results in:

$$\frac{d\psi_\tau}{d\varphi_\tau} = \frac{1}{4\pi^2K^2}\frac{2}{\sigma_0+\sigma_{0s}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\epsilon^{-(u'^2+v'^2)-u'v'[2\varphi_\tau/(\sigma_0+\sigma_{0s})]}\,du'\,dv'. \tag{16}$$

This integral can be evaluated by a method indicated by Rice by introducing two new variables y_1 and y_2 related to u' and v' by

$$u' = y_1 - \frac{a}{\sqrt{1-a^2}}y_2$$

$$v' = \frac{1}{\sqrt{1-a^2}}y_2$$

where

$$a = \frac{\varphi_\tau}{\sigma_0+\sigma_{0s}}.$$

Since

$$du'\,dv' = \frac{1}{\sqrt{1-a^2}}\,dy_1\,dy_2$$

and since the limits stay the same, the integral becomes

$$\frac{d\psi_\tau}{d\varphi_\tau} = \frac{1}{2\pi K^2}\frac{1}{\sigma_0+\sigma_{0s}}\frac{1}{\sqrt{1-a^2}}\cdot\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\epsilon^{-(y_1^2+y_2^2)}\,dy_1\,dy_2. \tag{17}$$

The double integral equals π and we obtain

$$\frac{d\psi_\tau}{d\varphi_\tau} = \frac{1}{2\pi K^2}\frac{1}{\sigma_0+\sigma_{0s}}\frac{1}{\sqrt{1-\left(\frac{\varphi_\tau}{\sigma_0+\sigma_{0s}}\right)^2}}. \tag{18}$$

By integration follows, with C as integration constant

$$\psi_\tau = \frac{1}{2\pi K^2}\sin^{-1}\left(\frac{\varphi_\tau}{\sigma_0+\sigma_{0s}}\right) + C. \tag{19}$$

The normalized input correlation function is

$$\rho_\tau = \frac{\varphi_\tau}{\sigma_0} \tag{20}$$

and with the parameter $\alpha = \sigma_{0s}/\sigma_0$ [as in (2)] ψ_τ becomes

$$\psi_\tau = \frac{\sin^{-1}\left(\frac{\rho_\tau}{1+\alpha}\right)}{2\pi K^2} + C. \tag{21}$$

ρ_τ equals one for $\tau = 0$, therefore

$$\psi_0 = \frac{\sin^{-1}\left(\frac{1}{1+\alpha}\right)}{2\pi K^2} \tag{22}$$

and the normalized output correlation function becomes

[4] S. O. Rice, "The mathematical analysis of random noise," *Bell Sys. Tech. J.*, vol. 23, pp. 282–332; 1944; vol. 24, pp. 46–156; 1945.
[5] G. E. Campbell and R. M. Foster, "Fourier Integrals for Practical Applications," Bell Sys. Monograph B–584, 1942. Pair 726.2 and 210.

$$\mathrm{P}_\tau = \frac{\psi_\tau}{\psi_0} = \frac{\sin^{-1}\left(\frac{\rho_\tau}{1+\alpha}\right)}{\sin^{-1}\left(\frac{1}{1+\alpha}\right)} \tag{23}$$

which is our final result.

That the integration constant C in (21) is zero may be shown in the following way.

For $\alpha = 1$ one would obtain, from (22)

$$\psi_0 = \frac{\sin^{-1}(\frac{1}{2})}{2\pi K^2} + C = \frac{1}{12K^2} + C. \tag{24}$$

Since ψ_0 represents the output rms value it may also be obtained by direct integration of the output probability distribution

$$\psi_0 = \int_{-1/2K}^{+1/2K} Ky^2\, dy = \frac{1}{12K^2}. \tag{25}$$

By comparison with (24) we find that indeed C must equal zero, as postulated in the previous discussion.

B. Evaluation of Output Spectrum for Gaussian Noise Passed Through an RC Filter

The output spectrum is given by (10).

$$S(\omega) = \frac{4}{\sin^{-1}\frac{1}{1+\alpha}} \int_0^\infty \sin^{-1}\left(\frac{\rho_\tau}{1+\alpha}\right) \cos(\omega\tau)\, \alpha\tau. \tag{26}$$

The solution is found by expansion of the $\sin^{-1}$ term in the integrand into a Taylor series and integrating term by term.

In particular for $\alpha = 1$

$$\sin^{-1}\left(\frac{\rho_\tau}{2}\right) = \frac{\rho_\tau}{2} + \frac{1}{2}\left(\frac{\rho_\tau}{2}\right)^3 + \frac{1.3}{2.4}\left(\frac{\rho_\tau}{2}\right)^5 + \cdots.$$

The series converges rapidly due to the factor (1/2) with which ρ_τ appears multiplied. Taking only the first term we obtain

$$S(\omega) = \frac{3}{\pi} 4 \int_0^\infty \rho_\tau \cos(\omega\tau)\, d\tau = \frac{3}{\pi} S_{\mathrm{in}}(\omega)$$

for any input spectrum $S_{\mathrm{in}}(w)$.

For ρ_τ according to (11) the output spectrum becomes

$$S(\omega) = \frac{3}{\pi}\frac{4}{B}\left[\frac{1}{1+\left(\frac{\omega}{B}\right)^2} + \frac{0.0417}{1+\left(\frac{\omega}{3B}\right)^2} + \frac{0.0047}{1+\left(\frac{\omega}{5B}\right)^2} + \cdots\right]$$

as compared with the normalized input spectrum

$$S_{\mathrm{in}}(\omega) = \frac{4}{B}\frac{1}{1+\left(\frac{\omega}{B}\right)^2}.$$

The factor $3/\pi$ equals 0.9549.

On the Role of Dynamic Programming in Statistical Communication Theory*

R. BELLMAN† AND R. KALABA†

Summary—In this paper we wish to show that the fundamental problem of determining the utility of a communication channel in conveying information can be interpreted as a problem within the framework of multistage decision processes of stochastic type, and as such may be treated by means of the theory of dynamic programming.

We shall begin by formulating some aspects of the general problem in terms of multistage decision processes, with brief descriptions of stochastic allocation processes and learning processes. Following this, as a simple example of the applicability of the techniques of dynamic programming, we shall discuss in detail a problem posed recently by Kelly. In this paper, it is shown by Kelly that under certain conditions, the rate of transmission, as defined by Shannon, can be obtained from a certain multistage decision process with an economic criterion. Here we shall complete Kelly's analysis in some essential points, using functional equation techniques, and considerably extend his results.

* Manuscript received by the PGIT, February 19, 1957.
† The RAND Corporation, Santa Monica, Calif.

I. Introduction

In this paper we wish to show that the fundamental problem of determining the utility of a communication channel in conveying information can be interpreted as a problem within the framework of multistage decision processes of stochastic type, and as such may be treated by means of the theory of dynamic programming.[1-3]

[1] R. Bellman, "Dynamic Programming," Princeton University Press, Princeton, N. J.; 1957.
[2] R. Bellman, "The theory of dynamic programming," *Bull. Amer. Math. Soc.*, vol. 60, pp. 503–515; November, 1954.
[3] R. Bellman, "Some functional equations in the theory of dynamic programming, I. Functions of points and point transformations," *Trans. Amer. Math. Soc.*, vol. 80, pp. 51–71; September, 1955.

This paper is to be envisaged as a step in the direction of a broad theory of communication, as contemplated by Wiener in his recent article,[4] and following the pioneering efforts of Rice,[5] and Shannon.[6] Among other steps along this path, we would like to cite the recent articles of Bussgang and Middleton,[7] and Middleton and Van Meter,[8] which employ the modern theory of statistical decision functions and sequential analysis, due to Wald.[9]

We shall begin by formulating some aspects of the general problem in terms of multistage decision processes, with brief descriptions of stochastic allocation processes and learning processes. Following this, as a simple example of the applicability of the techniques of dynamic programming, we shall discuss in detail a problem posed recently by Kelly.[10] In this paper, it is shown by Kelly that under certain conditions, the rate of transmission, as defined by Shannon[6] can be obtained from a certain multistage decision process with an economic criterion. Here we shall complete Kelly's analysis in some essential points, using functional equation techniques, and considerably extend his results.

In addition to the original problem of Kelly, we shall consider a time-dependent case, a process involving correlated signals, and a multisignal case, in both discrete and continuous versions. It will be seen that the logarithmic criterion function plays an extremely important role, as its special functional properties permit us to obtain explicit representations for both maximum return and optimal policy.

It is important to observe that the method we employ is equally applicable to criterion functions of varied analytic form. In any case it leads to a simple computational solution of the associated maximization problem. Furthermore, it is applicable to processes where the law of large numbers is not unrestrictedly applicable, such as those of finite duration, those with correlation, and so forth. A decision process automatically introduces a correlation between successive events.

Finally, we discuss briefly a functional equation arising from the general question of defining the "value" of a communication channel in a fashion which is independent of its use.

[4] N. Wiener, "What is information theory?" IRE Trans., vol. IT–2, p. 48; June, 1956.

[5] S. Rice, "Mathematical analysis of random noise," *Bell. Sys. Tech. J.*, vol. 23, pp. 282–332; July, 1944; vol. 24, pp. 46–156; January, 1945.

[6] C. Shannon, "A mathematical theory of communication," *Bell Sys. Tech. J.*, vol. 27, pp. 379–423; July, 1948; pp. 623–656; October, 1948.

[7] J. Bussgang and D. Middleton, "Optimum sequential detection of signals in noise," IRE Trans., vol. IT–1, pp. 5–18; December, 1955.

[8] D. Middleton and D. Van Meter, "Detection and extraction of signals in noise from the point of view of statistical decision theory," *J. Soc. Indust. Appl. Math.*, vol. 3, pp. 192–253; December, 1955; vol. 4, pp. 86–119; June, 1956.

[9] A. Wald, "Statistical Decision Functions," John Wiley & Sons, Inc., New York, N.Y.; 1950.

[10] J. Kelly, "A new interpretation of information rate," *Bell Sys. Tech. J.*, vol. 35, pp. 917–926; July, 1956.

II. The Underlying Model

Let us begin by constructing a simple model of one aspect of the general communication problem. It will be reasonably clear from what follows how more intricate models may be constructed to take account of more complicated systems.

Consider a source S which produces at discrete times[11] a sequence of pure signals, together with noise, which may be of either stochastic or deterministic type, depending upon our further assumptions concerning the structure of the system. The combined signal is fed into a black box which we call a "communication channel," which, in turn, emits a signal. This output signal is observed.

On the basis of the observation of the output signal, it is desirable to make various deductions concerning the properties of the original pure signal.

Schematically,

In mathematical terms, let

x = the pure signal emanating from S.
r = the noise associated with the signal.
$x' = F(x, r)$, the input to the communication system.
y = the signal transmitted to the observer by the communication channel. (1)

Let us further write

$$y = T(x') = T(F(x, r)), \tag{2}$$

where T represents the transformation of the input signal x' due to the communication channel.

Consider the set of all communication systems, or, equivalently, the space of all associated transformations, T. We wish to introduce an ordering, or, what is much more satisfactory when possible, a metric which will enable us to compare two communication systems and to evaluate their performance (Section X-C).

It must be stressed at the very outset of an investigation of this type that it should be possible to accomplish this aim in an unlimited number of ways, dependent upon the source, channel, nature of the observer, and the personal philosophies involved; *i.e.*, upon the utility scales employed.

In the following sections, we shall present two alternate methods for evaluating a communication system. Although each is a particular case of a more general scheme, which we shall discuss subsequently, it is worthwhile to present them separately first, as they occur in important appli-

[11] The case of continuous signal emission can also be treated by the methods outlined below, at the expense of the introduction of more sophisticated concepts. We prefer to keep the mathematical level moderate in this first discussion; however, see Section X-D

tions. In this way, we hope to avoid the usual risk of bscuring the issue by extreme generality.

III. A Stochastic Allocation Process

Let us assume that the observer has a sum of money, r resources of other types, which we denote by the vector called the state vector. Upon receiving a y signal, the bserver is required to make an allocation of resources various activities. The effect of this allocation is to hange x into $R(x, y)$, a stochastic vector whose distriution we shall assume here to be known. The case in hich the distribution is not known is closely allied with e second model we shall discuss.

The process is now repeated N times, where N may be xed, which is the simplest case, or the number of stages ay depend upon the process itself as a consequence of preassigned stop rule. Let us again consider only the mplest case, that of fixed N.

Further, let us suppose that the purpose of the observer carrying out this process is to maximize the expected alue of some function of his final state vector, the state ttained after N stages of the process.

Let f_T denote this maximum expected value, and f_I the aximum expected value when the transformation T is e identity transformation, the case in which we have a istortionless communication channel.

Let us then agree to measure the worth of the original ommunication system by means of a preassigned function f f_T and f_I. In this fashion we introduce a metric into the pace of transformations T, and thus, into the set of ommunication channels. The simplest cases are those here we use a function of $f_I - f_T$, or a function of f_T/f_I.

We shall discuss a simple case of a process of the above ype in later sections. For the formulation and mathematical discussion of some particular processes of this eneral type we refer to Robbins[12] and Bellman.[13]

IV. A Stochastic Learning Process

Let us now consider a different type of stochastic rocess. The observer is required to make a decision conerning the nature of the pure signal emitted by S. ubject to constraints imposed by the costs of observation nd by limitations of time, he can observe as many amples of the signal emitted by the communication ystem as he wishes.

As a result of these decisions, he makes an estimate oncerning properties of the pure signal, and thereby inurs a cost dependent upon the deviation of this estimate rom the actual situation.

The problem is to carry out the process of first obserration and then estimation so as to minimize the expected total cost, where the total cost is a given function of the costs of observation and the cost of deviation.

The theory of sequential analysis is devoted to one aspect of this general problem. Other aspects arise in the theory of learning processes.[10,13]

It is clear that we can define the worth of the communication channel in a manner completely analogous to the procedure discussed above.

V. A More General Process

It is clear that both processes are particular cases of a more general process where neither the structure of, nor the transformation due to, the communication channel is completely known. Each stage of the process yields a certain return, which may be negative, in resources, and yields additional information, which may also be negative, concerning the intrinsic structure of the combined system.

The problem is to carry out the sequence of decisions so as to maximize some function, which may not necessarily be completely known, of the total returns and the information pattern.

It is interesting to observe that posed in this way, we encounter one of the basic problems of experimental research.

VI. Discussion

For the above approach to be fruitful and to represent more technology than tautology, one must possess mathematical techniques capable of formulating in precise terms, and of treating, processes of the kind described above.

The theory of sequential analysis developed by Wald, Wolfowitz, Blackwell, and Girshick provides an approach to one class of problems of this type, while a general approach to these multistage decision processes is provided by the theory of dynamic programming of Bellman.[1-3]

As an application of these general methods, we shall consider a simple interesting model proposed recently by Kelly,[10] and some generalizations.

VII. The Model of Kelly

Let us begin by treating the first problem posed by Kelly.

A gambler receives advance information concerning the outcomes of a sequence of independent sporting events over a noisy communication channel. We assume that the outcome of each event is the result of play between two evenly matched teams, and that p is the probability of a correct transmission, and $q = (1 - p)$ the probability of an incorrect transmission.

Assuming that the gambler starts with an initial amount x and bets on the outcome of each event so as to maximize his expected capital at the end of N stages of play, it is clear that he wagers his entire fortune on each play if $p > \frac{1}{2}$, and nothing if $p < \frac{1}{2}$. If $p = \frac{1}{2}$, it makes no difference what policy he employs. (We are supposing that the gambler must bet on the received signal, if at all. It

[12] H. Robbins, "Some aspects of the sequential design of experiments," *Bull. Amer. Math. Soc.*, vol. 58, pp. 527–536; September, 1952.

[13] R. Bellman, "A problem in sequential design of experiments," *Sankhya*, vol. 16, pp. 221–229; April, 1956.

is easy to see that if we allow him complete freedom in placing bets, then, in the case where $p < \frac{1}{2}$, his bet will always be contrary to the information he receives.) If $\frac{1}{2} < p < 1$, with probability one the gambler will go broke following such a policy, since eventually he must lose a bet.

A much more difficult process arises if we take p to be a fixed, but unknown quantity which must be determined on the basis of the observed results of betting. This leads to a "learning process." An expository treatment containing a number of additional references may be found in Robbins,[12] while a treatment by dynamic programming of a similar problem may be found in Bellman.[13]

Let us now assume that the above mode of play appears too hazardous to the gambler, and that he wishes to pursue a more conservative policy, one that will prevent him from ever being wiped out. He may then proceed to maximize the expected value of the logarithm of his capital at the end of N stages of play (see Section XI).

For the one-stage process, he is faced with the problem of maximizing

$$E_1(y) = p \log (x + y) + q \log (x - y) \tag{3}$$

over all y in $[0, x]$. Here y is the amount wagered, even odds being assumed. It is easy to see that, if $p > q$, we have as the maximizing value of y,

$$y = (p - q)x, \tag{4}$$

and for that value of y

$$E_1 = \log x + \log 2 + p \log p + q \log q. \tag{5}$$

If $p \leq q$, the maximum is at $y = 0$.

It is not difficult to show that if we consider N-stage processes. where we restrict ourselves to wagering policies which require the wagered amount to be a *fixed* proportion of the total capital at each stage, then the policy described above is optimal. This result was established by Kelly[10] in a very ingenious fashion.

Let us now demonstrate that this policy is optimal within the class of all wagering policies.

VIII. Economic Forecasting

It is clear that the above mathematical model is abstractly identical with problems that arise in connection with economic forecasting, in particular, and with forecasting, in general, as for example, weather prediction.

In these cases, the physical world is the source and the scientific corps, both experimental and theoretical, the communication channel. Sometimes, the theorist or experimenter is also the observer; at other times, it is the business man or politician who must decide to what extent he trusts his communication channel.

IX. Dynamic Programming Approach[14]

Let us begin by formulating the problem in dynamic programming terms. Define the following sequence of functions,

$f_N(x)$ = expected value of the logarithm of the final capital obtained from an N-stage process starting with an initial capital x and using an optimal policy. (6)

An *optimal policy* is here defined as one which maximizes the expected value of the logarithm of the capital at the end of N stages.

The mathematical results we shall obtain depend upon the intuitive *principle of optimality: An optimal policy has the property that, whatever the initial state and initial decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision.*[2]

For this particular process, the meaning of the statement is that, regardless of how much is wagered by the gambler at the first stage, and regardless of whether he wins or loses, he utilizes the amount of money in his possession prior to the betting at the second stage in such a way as to maximize; *i.e.*, he proceeds so as to maximize the expected value of the logarithm of his capital after the remaining (N-1) stages.

The mathematical transliteration of this statement yields the recurrence relations

$$\begin{aligned} f_1(x) &= \log x + K, \\ f_N(x) &= \max_{0 \leq y \leq x} [p f_{N-1}(x + y) \\ &\qquad + (1 - p) f_{N-1}(x - y)], \qquad N \geq 2, \end{aligned} \tag{7}$$

where

$$K = \begin{cases} \log 2 + p \log p + q \log q, & p > q, \\ 0, & p \leq q. \end{cases} \tag{8}$$

Let us now demonstrate the

Theorem: For $N \geq 1$, *we have*

$$f_N(x) = \log x + NK, \tag{9}$$

where K is defined as above. The optimal policy is unique and independent of N. It consists of choosing

$$y = (p - q)x, \qquad p > q, \tag{10a}$$

$$y = 0, \qquad p \leq q. \tag{10b}$$

Proof. Let us proceed inductively, beginning with the known result for $N = 1$. Assuming that the result holds for N, we have,

$$\begin{aligned} f_{N+1}(x) &= \max_{0 \leq y \leq x} [p [\log (x + y) \\ &\qquad + NK] + (1 - p) [\log (x - y) + NK]] \\ &= \max_{0 \leq y \leq x} [p \log (x + y) \\ &\qquad + (1 - p) \log (x - y)] + NK. \end{aligned} \tag{11}$$

[14] The results contained in this section answer the fundamental question posed by Kelly, *op. cit.*, p. 926.

$$f_{N+1}(x) = (\log x + K) + NK = \log x + (N+1)K. \quad (12)$$

The statement concerning the form of the optimal policy follows from the analytic form of $f_N(x)$.

Now that the "best" performance of the noisy channel has been determined, it may be compared in various possible ways with the performance of a perfect channel.

X. Generalizations

A. *Time Dependent Case*

Before proceeding to more general cases, let us consider a simple extension of the above model.

First, let us suppose that at the kth stage the probability of correct transmission is p_k, and of incorrect transmission $q_k = 1 - p_k$. For fixed N, define the sequence of functions

$f_k(x)$ = expected value of the logarithm of the final capital obtained from the remaining k stages of the original N-stage process, starting with an initial capital x, and using an optimal policy.

Then

$$f_1(x) = \underset{0 \le y \le x}{\text{Max}} \{p_N \log (x + y) + q_N \log (x - y)\},$$

$$f_k(x) = \underset{0 \le y \le x}{\text{Max}} \{p_{N-k+1}\ f_{k-1}\ (x + y) + q_{N-k+1}\ f_{k-1}\ (x - y)\}, \qquad N \ge k \ge 2. \quad (14)$$

As before, it follows inductively that

$$f_k(x) = \log x + k \log 2 + \sum_{r=N-k+1}^{N} [p_r \log p_r + q_r \log q_r], \quad (15)$$

provided that $p_k > \frac{1}{2}$ for $k = 1, 2, \cdots, N$. Whenever this condition fails, the term $p_k \log p_k + q_k \log q_k$ must be replaced by $(-\log 2)$.

B. *Correlation*

Let us now consider the case where the signals are not independent. The simplest case, perhaps, is that where the probability of correct transmission p_k depends upon whether or not the preceding signal was transmitted correctly. Although a large variety of questions of this type may be formulated, we feel that the following discussion will illustrate the uniform method which may be employed to treat them.

Let

p_k = probability of correct transmission of the kth signal, if the $(k-1)$st signal was transmitted correctly.

r_k = probability of correct transmission of the kth signal, if the $(k-1)$st signal was transmitted incorrectly. (16)

Define the sequence of functions,

$f_k(x)$ = expected value of the logarithm of the final capital obtained from the remaining k stages of the original N-stage process, starting with an initial capital x, and the information that the $(k-1)$st signal was transmitted correctly, using an optimal policy.

$g_k(x)$ = the corresponding function in the case where the $(k-1)$st signal was transmitted incorrectly. (17)

Then, as above,

$$f_k(x) = \underset{0 \le y \le x}{\text{Max}} [p_{N-k+1} f_{k-1}(x + y) + (1 - p_{N-k+1}) g_{k-1}(x - y)]$$

$$g_k(x) = \underset{0 \le y \le x}{\text{Max}} [r_{N-k+1} f_{k-1}(x + y) + (1 - r_{N-k+1}) g_{k-1}(x - y)]. \quad (18)$$

It follows inductively, as before, that

$$f_k(x) = \log x + a_k,$$
$$g_k(x) = \log x + b_k, \quad (19)$$

where the recurrence relations for the a_k and b_k are readily established.

C. *Multiple Signal Channels*

Let us now consider the situation in which the channel is called upon to transmit any of M different signals, with the probability of correct transmission dependent upon the signal transmitted.

Assume that the gambler possesses the following information:

p_{ij} = the conditional probability that the j signal was sent if the i signal was received.

q_i = the probability of receiving the i signal at any time.

r_j = the return from a unit winning bet on the j signal. (20)

For simplicity of notation principally, we assume that there is no time dependence. It will be readily seen from what has preceded and what follows that this factor could easily be included.

The process proceeds as follows. Upon receiving a signal, say the i signal, the gambler is free to bet an amount z_j that the signal actually sent was a j signal. The restrictions upon z_j are

$$z_j \ge 0 \quad (21a)$$

$$\sum_{j=1}^{M} z_j \le x \quad (21b)$$

where x is the current quantity of capital.

As before, we assume that the purpose of the process is to maximize the expected value of the logarithm of the capital remaining after N stages. Let us then define

the function $f_N(x)$, for $N = 1, 2, \cdots, x \geq 0$, as the expected value of the logarithm, starting with a capital of x, and using an optimal policy.

Observing that there is a probability p_{ij} of winning $r_j z_j$ and losing on the other bets when the i signal is received, we see that with probability p_{ij} the gambler has the quantity $r_j z_j + x - \sum_{s=1}^{M} z_s$ left after one stage when the i signal is received.

Consequently, employing the principle of optimality as above, we obtain the recurrence relations

$$f_N(x) = \sum_{i=1}^{M} q_i \left\{ \operatorname*{Max}_{R(x)} \sum_{j=1}^{M} p_{ij} \cdot f_{N-1}\left(r_j z_j + x - \sum_{s=1}^{M} z_s \right) \right\}, \qquad N \geq 2$$

$$f_1(x) = \sum_{i=1}^{M} q_i \left\{ \operatorname*{Max}_{R(x)} \sum_{j=1}^{M} p_{ij} \cdot \log \left(r_j z_j + x - \sum_{s=1}^{M} z_s \right) \right\}. \tag{22}$$

Here $R(x)$ is the region in z_j space defined by (21).

Let us now prove inductively that

$$f_N(x) = \log x + NK, \tag{23}$$

where K is a constant independent of x given by

$$K = \sum_{i=1}^{M} q_i \left\{ \operatorname*{Max}_{R(1)} \sum_{j=1}^{M} p_{ij} \log \left(r_j z_j + 1 - \sum_{s=1}^{M} z_s \right) \right\}. \tag{24}$$

The result is true for $N = 1$, as we see upon setting $N = 1$ in (23) and comparing with (22). Assume that the result holds for $N - 1$ and substitute the expression for $f_{N-1}(x)$ given in (23) in (22). Collecting terms, we obtain the expression in (23) for N.

From the expression for K it is clear that the optimal policy depends only on p_{ij} and r_i and not on the q_i, though the return itself does also depend on q_i. An interesting special case is that in which it is required that $\sum_{1=1}^{M} z_i = x$; that is, the gambler is required to wager all of his available funds. In this case the optimal policy depends only on (p_{ij}); *i.e.*, on the communication channel, and not at all on q_i or on r_i.

Schematically

$$i \longrightarrow \boxed{t_{ij}} \longrightarrow j.$$

If we now introduce the quantities

p_i = probability of sending an i,

t_{ij} = the conditional probability that if an i is sent, then j is received,

we then have

$$p_i = \sum_{j=1}^{M} q_j p_{ji}$$

$$q_i = \sum_{j=1}^{M} p_j t_{ji},$$

and (t_{ij}) is the inverse of (p_{ij}). Consequently, in this case the *optimal policy* is dependent only on (t_{ij}), which characterizes the communication channel and is independent of both the source characterized by p_i and the outside world, characterized by the odds, r_i. The *return*, however, does depend on all these quantities.

These considerations are significant, for they imply that the gambler's actions are controlled solely by the quality of the communication channel, though his ultimate return is determined by the situation *in toto*. This leads to the possibility of comparing two channels under the same conditions or of evaluating the performance of a given channel under various conditions.

Specializations to the unsymmetric binary channel are immediate.

D. Continuum of Signals

Consider now the case where there is a continuum of different signals. Let

$dG(u, v)$ = the conditional probability that a signal with label between v and $v + dv$ is sent if the u signal is received, $-\infty < u, v < \infty$, (25)

and let

$dH(u)$ = the probability that a signal with label between u and $u + du$ is received at any stage. (26)

Then, considering the process corresponding to the special case discussed above, even bets being assumed for the sake of simplicity, we derive the recurrence relations

$$f_N(x) = \int_{-\infty}^{\infty} \left[\operatorname*{Max}_{z(v)} \int_{-\infty}^{\infty} f_{N-1}(2z(v))\, dG(u, v) \right] dH(u),$$

$$f_1(x) = \int_{-\infty}^{\infty} \left[\operatorname*{Max}_{z(v)} \int_{-\infty}^{\infty} \log (2z(v))\, dG(u, v) \right] dH(u). \tag{27}$$

In both cases, the maximization is over all functions $z(v)$ satisfying the conditions

$$z(v) \geq 0 \tag{28a}$$

$$\int_{-\infty}^{\infty} z(v)\, dv = x. \tag{28b}$$

As above, it is easily seen inductively that

$$f_N(x) = \log 2x + KN, \tag{29}$$

where

$$K = \int_{-\infty}^{\infty} \operatorname*{Max}_{z(v)} \left[\int_{-\infty}^{\infty} \log z(v)\, dG(u, v) \right] dH(u), \tag{30}$$

and

$$z(v) \geq 0, \tag{31a}$$

$$\int_{-\infty}^{\infty} z(v)\, dv = 1. \tag{31b}$$

XI. Criterion Functions Yielding Invariant Policies

We have seen above that the linear function yields an variant policy at each stage, and likewise the logarithm. is of interest to determine all criterion functions possing this property.

The following version of the problem will be treated re. Let $\phi(x)$ be a monotone increasing concave function fined over $0 < x \leq 1$. Consider the one-stage process here we wish to maximize

$$E(y) = p\phi(x + y) + (1 - p)\phi(x - y). \tag{32}$$

he function $E(y)$ is concave as a function of y for $0 \leq y$ x, $0 < x \leq 1$, and thus has a unique maximum, unless $x)$ is linear and $p = \frac{1}{2}$. Let us dismiss the case of linrity by requiring strict concavity, $\phi''(x) < 0$, and take $> \frac{1}{2}$.

Let us assume that, for all x in $0 < x \leq 1$, there is a lution of

$$\frac{dE}{dy} = p\phi'(x + y) - (1 - p)\phi'(x - y) = 0 \tag{33}$$

aving the form

$$y = r(p)x, \tag{34}$$

here $r(p)$ is a nonnegative differentiable function of p r $\frac{1}{2} < p \leq 1$, possessing a continuous derivative.

Then (33) is equivalent to the functional equation

$$\phi'(x(1 + r(p))) = (1 - p)\phi'(x(1 - r(p))),$$
$$\text{for} \quad 0 < x \leq 1, \qquad \tfrac{1}{2} < p \leq 1. \tag{35}$$

Let $x(1 + r(p)) = y$. Then (35) reduces to

$$\frac{p}{1 - p}\,\phi'(y) = \phi'\left(\frac{y(1 - r(p))}{(1 + r(p))}\right). \tag{36}$$

We now differentiate first with respect to y, and then ith respect to p, obtaining the two equations

$$\frac{p}{-p}\,\phi''(y) = \left(\frac{1 - r(p)}{1 + r(p)}\right)\phi''\left(\frac{y(1 - r(p))}{(1 + r(p))}\right),$$
$$\frac{1}{(1 - p)^2}\,\phi'(y) = y\,\frac{d}{dp}\left(\frac{1 - r(p)}{1 + r(p)}\right)\phi''\left(\frac{y(1 - r(p))}{1 + r(p)}\right). \tag{37}$$

Dividing the two equations, we obtain

$$\frac{y\phi''(y)}{\phi'(y)} = \frac{u(p)}{p(1 - p)(du/dp)}, \tag{38}$$

where $u(p) = (1\text{-}r(p))/(1 + r(p))$.

Since the left side is a function of y and the right side a function of p, both sides must be constant. Setting

$$\frac{y\phi''(y)}{\phi'(y)} = K, \qquad K \leq 0, \tag{39}$$

we obtain

$$\log \phi'(y) = K \log y + c_1. \tag{40}$$

Hence

$$\phi'(y) = c_2 y^K. \tag{41}$$

Without loss of generality, let us normalize, so that $\phi'(1) = 1$. Then

$$\phi'(y) = y^K. \tag{42}$$

If $K > -1$, we have

$$\phi(y) = \frac{y^{K+1}}{K + 1} + c_1'. \tag{43}$$

If $K = -1$, we have

$$\phi(y) = \log y + c'. \tag{44}$$

It is clear that $K \geq -1$ is necessary for $\phi(y)$ to be nonnegative for $y > 0$. Finally, without loss of generality, we can let $c' = c_1' = 0$.

XII. Discussion

In the foregoing pages, we have essayed to describe some applications of the concepts and techniques of the theory of dynamic programming to various aspects of communication theory. As simple illustrations we have considered a particular process discussed by Kelly and various generalizations. In subsequent papers, we propose to treat in greater detail some mathematical models of greater scope.

Complex Processes for Envelopes of Normal Noise*

RICHARD ARENS†

Summary—The paper presents a brief exposition of the technique of complex normal random variables as utilized in the study of the envelopes of Gaussian noise processes. The central concept is the pre-envelope $z(\ )$ of a real normal process. The pre-envelope $z(\ )$ of a real function $x(\ )$ is a complex function whose real part is $x(\ )$ and whose absolute value is the envelope, in the sense of high-frequency theory, of $x(\ )$. The joint probability density for $z(t)$, $z'(t)$ is found and used to get the threshold crossing rate. Consideration of nonstationary processes is included.

I. Introduction

THE envelope $r(t)$ of a real function x of one variable ("time") is obtained, according to one conception[1,2] of the envelope, by forming the conjugate (Hilbert transform) function y of x and taking the absolute value $|z(t)|$ of $z(t) = x(t) + iy(t)$. We call the function z the *pre-envelope* of x. Its spectrum is supported wholly by one half the frequency axis; we choose the positive half. When two "signals" x_1, x_2 are superposed, the pre-envelope of $x_1 + x_2$ is the sum of the pre-envelopes z_1, z_2 of x_1, x_2, respectively; and more generally if x is filtered with output at time t

$$\int_0^\infty W(t - s)x(s)\, ds \tag{1}$$

then the pre-envelope of the output is at time t

$$\int_0^\infty W(t - s)z(s)\, ds \tag{2}$$

where z is the pre-envelope of x. In dealing with linear systems it is evidently better to utilize the pre-envelope, because then at each stage between successive filterings the envelope can be readily inspected. (This procedure is often employed for aesthetic reasons independent of envelopes.) We shall, however, be concerned with random processes rather than a single time series in this paper.

Accordingly, let x be a sample function from some real normal (so-called "Gaussian") stationary process. Form $z = x + iy$ where z is the pre-envelope of x. This gives rise to a new random process which is stationary, normal and, of course, complex. We first prove this and then illustrate the utility of such complex processes by calculating the "alarm rate" (see Section V) or threshold-crossing rate for a "noise" envelope. Our object is, however, more to exhibit the technique and illustrate the principles of complex valued random processes. Our experience ha been that such technique is practically indispensable fo treating envelope questions for nonstationary processe and considerably more convenient for arriving at th known results than the known treatment of the stationar case.

* Manuscript received by the PGIT, February 26, 1957.
† Dept. of Math., Univ. of Calif., and RCA, Los Angeles, Calif.
[1] S. O. Rice, "Mathematical analysis of random noise," *Bell. Sys. Tech. J.*, vol. 23, p. 282; 1944, and vol. 24, p. 109; 1945.
[2] V. I. Bunimovitch, "The fluctuation process as a vibration with random amplitude and phase," *J. Tech. Phys., USSR*, vol. 14, p. 1231; November, 1949.

II. The Pre-Envelope Process Examined

Bunimovich[3] has pointed out that if we have a rea valued function,

$$x(t) = \int_{-\infty}^{\infty} \exp(2\pi i f t)\, dX(f), \quad (3$$

where X is of bounded variation and $dX(-f) = \overline{dX(f}$ since x is real, then the absolute value of

$$x(t) + iy(t) = z(t)$$
$$= 2\int_{0+}^{\infty} \exp(2\pi i f t)\, dX(f) + \Delta X \Big|_{0-}^{0+} \quad (4$$

is the Rice[4] envelope of (3). It is clear that z is *linearl* determined by x. Hence z is also[5] a normal, stationar process, although, as pointed out earlier, a *complex* one Here we confine ourselves to the calculation of the *auto correlation* for the z process, which is by definition (E = *expected value* or *expectation*),

$$\tfrac{1}{2}E[z(t)\overline{z(s)}] = R_c(t - s), \quad (5$$

a justified notation because it depends only on $t - s$. W aim to express $R_c(t)$ in terms of the integrated powe spectrum[6] W for the x process, in terms of which th autocorrelation of the latter is given by

$$E[x(t)x(s)] = R(t - s) = \int_{0-}^{\infty} \cos 2\pi f(t - s)\, dW(f). \quad (6$$

We expand the cosine, obtaining briefly an array *cos co* + *sin sin*, where the first factor in each case depends on and the second on s. For the expectation of $x(t)y(s)$ w remark that the expectation of a linear functional is th linear functional of the expectation.[7] The linear functiona

[3] *Ibid.*
[4] Rice, *op. cit.* The author gratefully acknowledges that this wa pointed out by Prof. J. Dugundji.
[5] J. L. Doob, "Stochastic Processes," John Wiley & Sons, Inc New York, N.Y.; 1953 and A. Khintchine, "Korrelationstheor der stationären stochastischen Prozesse," *Math. Ann.*, vol. 10 p. 604; 1934.
[6] Rice, *op. cit.*
[7] In practice this reduces to interchanging the order of sever limit-operations. In what follows we do not trouble to justify th interchange of operations, because our purpose is mainly expositio (see footnote 10).

re intended is the Hilbert transform applied to the ond appearing function x, in (6). Thus the expectation $x(t)y(s)$ is an integral like (6) with *cos cos* + *sin sin* placed by *cos sin* − *sin cos*, since the Hilbert transform *cos* is *sin* and that of *sin* is −*cos*. Operating, instead, the t-dependent factors gives the expectation of $y(t)$ $s)$ with the array *sin cos* + *cos sin*, and for $y(t)y(s)$ we tain the array *sin sin cos cos*. Adding the integrals for $t)x(s)$, $y(t)y(s)$, $-ix(t)y(s)$, and $iy(t)x(s)$, and dividing 2 gives the quantity called for in (5), whence

$$R_c(t - s) = \int_{0-}^{\infty} \exp\,(2\pi i f(t - s))\,dW(f). \tag{7}$$

III. Complex Normal Distributions

Just as in vector analysis, there is a point in introducing mplex processes only if in subsequent operations, no rther reference to the real components is necessary. order to gain such facility, some basic ideas have to be pt in mind. The probability density $F(\ )$ of a complex ndom variable z is a real function defined in the complex ane. Also it is nonnegative and has integral one over e plane. If, for example,

$$F(z) = \frac{1}{2\pi\sigma^2} e^{-z\bar{z}/2\sigma^2} \tag{8}$$

en we call z a normally distributed random variable ith zero mean and semivariance σ^2. This latter is half e mean (E) of $z\bar{z} = |z|^2$. The probability density for $= |z|$ can be obtained from (8) by changing to polar ordinates and integrating over the angle and takes the rm

$$F_1(r) = r\sigma^{-2} \exp\,(-2^{-1}\sigma^{-2}r^2) \qquad (r \geq 0), \tag{9}$$

metimes called "Rayleigh type with parameter σ^2." w is a complex normally distributed random variable ith mean m which may, of course, be complex, then $= m + z$ where z is distributed by (8). The probability ensity for $r = |w|$ can be obtained[8] via polar coordinates, nd is

$$F_2(r) = r\sigma^{-2} I_0(r\sigma^{-2}\mu) \exp\,[-2^{-1}\sigma^{-2}(r^2 + \mu^2)] \tag{10}$$

here $r \geq 0$, and $\mu = |m|$. This corresponds to the distribution of the envelope of signal plus noise. μ is the gnal envelope and σ^2 is the mean square of the noise.

The typical jointly normal probability density for n omplex random variables is given by

$$F(z_1, \cdots, z_n) = C \exp\left(-2^{-1} \sum_{i,j=1}^{n} q_{ij} z_i \bar{z}_j\right) \tag{11}$$

here $q_{ij} = \bar{q}_{ji}$ and $C = (2\pi)^{-n} \times$ determinant of (q_{ij}).

The inverse (r_{ij}) of the matrix (q_{ij}) has this property[9]

$$r_{ij} = \tfrac{1}{2} E[\bar{z}_i z_j]. \tag{12}$$

IV. Return to the Pre-Envelope Process

In a possibly nonstationary, normal process we have a joint probability density like (11) for the values $z(t_1) \equiv z_1, \cdots, z(t_n) \equiv z_n$ of a sample function z, and the r_{ij} of (12) takes the form

$$r(t_i, t_j) = \tfrac{1}{2} E[\overline{z(t_i)} z(t_j)] \tag{13}$$

where r is a function of two variables characterizing the process. If the process is *stationary*, this depends only on $t_i - t_j$ and may be written $r_{ij} = r(t_i, t_j) = R_c(t_j - t_i)$. In view of (5), this R_c is given for the pre-envelope process by (7).

If a process characterized by (13) (*i.e.*, by specifying r) is subjected to a *linear* operation $w = L(z)$, then w is again a normally distributed complex random variable, provided there are no[10] convergence difficulties in the carrying out of the operation L. In carrying out the operation L on the function z, there will usually occur a *dummy variable t*, as in the examples

$$L(z) = a\left[\frac{d}{dt} z(t)\right]_{t=t_1} \tag{14}$$

and

$$L(z) = \int_a^b A(t) z(t)\,dt. \tag{15}$$

These and similar functional operations may be written as

$$L(z) = L_{(t)}[z(t)]. \tag{16}$$

Naturally, the t in (14)-(16) may be replaced by any other letter without altering the significance, hence the name *dummy variable*. Supposing that the questions alluded to[10] do not impede us, then[11]

[8] J. L. Lawson and G. E. Uhlenbeck, "Threshold Signals," I.I.T. Rad. Lab. Ser., McGraw-Hill Book Co., Inc., New York, .Y., vol. 24, p. 154; 1947.

[9] These facts can be established by starting with a *characteristic function*

$$\varphi(\zeta_1, \cdots, \zeta_n) = \exp\,(-2^{-1}Q)$$

where $Q = \Sigma r_{ij}\bar{\zeta}_i\zeta_j (r_{ij} = \bar{r}_{ji})$. The Fourier relation between φ and F is that

$$\varphi(\zeta_1, \cdots, \zeta_n) = \int F(z) \exp iB\,dx_1 \cdots dy_n$$

where $B = \Sigma(x_k\xi_k + y_k\eta_k)$, $z_k = x_k + iy_k$, $\zeta_k = \xi_k + i\eta_k$, and the integration is over R^{2n}. The details are easily adapted from the real variable case as handled in H. Cramér, "Mathematical Methods of Statistics," Princeton University Press, Princeton, N. J.; 1946.

[10] This seems to be the case in applications. We refer to Doob, *op. cit.*, Khintchine, *op. cit.*, V. Cramér, "On the theory of stationary random processes," *Ann. Math.*, vol. 41, p. 215; 1940, and A. Blanc-Lapierre and R. Fortet, "Theorie des Fonctions Aleatoires," Masson, Paris; 1953, for the mathematical theory.

[11] A bar placed over the L in our examples means that the complex coefficient a in (9) and the function A in (10) have to be replaced by their conjugates.

$$\tfrac{1}{2}E[|w|^2] = \tfrac{1}{2}E[\overline{L_{(t)}[z(t)]}L_{(s)}[z(s)]]$$
$$= \overline{L_{(t)}}L_{(s)}[\tfrac{1}{2}E[\overline{z(t)}z(s)]]$$
$$= \overline{L_{(t)}}L_{(s)}[r(t, s)].$$

Thus the probability density of $w = L(z)$ *is given by*

$$(2\pi)^{-1}\tau^{-2} \exp[-2^{-1}\tau^{-2} \mid w \mid^2] \quad (17)$$

where τ^2 *is given by*

$$\tau^2(=\tfrac{1}{2}E[\mid w \mid^2]) = \overline{L}_{(t)}L_{(s)}[r(t, s)]. \quad (18)$$

The full utility of the complex method (examples in Section V) is not attained until also two or more linear functions of a given process z can be treated. Such a generalization of (18) is easily obtained. Let the linear functionals be

$$w_1 = L_1[z], \cdots, w_n = L_n[z]. \quad (19)$$

Then, analogous to, and generalizing (18),

$$\tfrac{1}{2}E[\bar{w}_i w_j] = \overline{L_{i(t)}}L_{j(s)}[r(t, s)]. \quad (20)$$

If these numbers are taken as the r_{ij} *and the* q_{ij} *calculated in terms of them, then* (11) *gives the joint probability density of the variables* (19). *Of course the* z'*s in* (11) *have to be replaced by* w'*s.*

V. Alarm Rate for the Envelope

By this we mean the following. Let x be a stationary normal process specified by an R as in (6). Let $P(T)$ be the probability that the envelope will rise above a level V in a certain interval of time T, *i.e.*,

$$P(T) = \text{prob. } \{\mid z(0) \mid < V < \mid z(T) \mid\}. \quad (21)$$

If for small T we have

$$P(T) = pT + q$$

where q tends to 0 faster than T, as T tends to 0, then p is the *alarm rate* associated with the level V. Now (21) could be evaluated by setting up the bivariate distribution (11) again mentioned in Section V and integrating, but this labor is unnecessary.

Rice[1] solves this problem by using the joint distribution for $\mid z(t)\mid$ and $d/dt \mid z(t)\mid$. We shall, however, use the joint distribution of the two linear functionals

$$w_1 = L[z] = [z(t)]_{t=0},$$
$$w_2 = M_{(t)}[z(t)] = [z'(t)]_{t=0} \quad (22)$$

and hope that our performance may encourage the reader to use complex variables for other envelope problems.

We actually do not need stationarity; so let the z process be specified as by (13). [We recapitulate: if the datum *is* a real process such as (6), we form first (7) and take $r(t, s) = R_c(s - t)$.] With the r_{ij} to use in (11) for $w_1, w_2 (i, j = 1, 2)$ we have, according to (20), the following array:

$$\begin{matrix} [r(t, s)]_{s=t=0} & \left[\dfrac{\partial}{\partial s} r(t, s)\right]_{s=t=0} \\ \left[\dfrac{\partial}{\partial t} r(t, s)\right]_{s=t=0} & \left[\dfrac{\partial^2}{\partial s\, \partial t} r(t, s)\right]_{s=t=0} \end{matrix}$$

i.e., we have the automatically hermitean matrix

$$\begin{matrix} r(0, 0) & r_2(0, 0) \\ r_1(0, 0) & r_{12}(0, 0) \end{matrix} \quad (23)$$

where the indices in (23) denote partial derivatives. Let the determinant of (23) be called D. Then the *inverse* is, omitting the (0, 0)'s from (23),

$$(q_{ij}) = \frac{1}{D}\begin{pmatrix} r_{12} & -r_2 \\ -r_1 & r \end{pmatrix}. \quad (24)$$

The joint probability density for $z(0)$, $z'(0)$ is, therefore

$$F(z, w) = D(2\pi)^{-2} \exp(-2^{-1}Q) \quad (25)$$

where $Q = D^{-1}(r_{12}\mid z \mid^2 - 2\mathbf{R}r_1\bar{w}z + r \mid w \mid^2)$, $\mathbf{R}$ denoting "real part."

In the *stationary case*, (23) takes the form

$$\begin{matrix} R_c(0) & +R_c'(0) \\ -R_c'(0) & -R_c''(0). \end{matrix} \quad (26)$$

Reference to (7) discloses that

$$R_c(0) = \int_{0-}^{\infty} dW(f) = R(0) = \text{"mean power"}$$

$$R_c'(0) = 2\pi i \int_{0-}^{\infty} f\, dW(f) \quad \text{(pure imaginary)}$$

and

$$R_c''(0) = -4\pi^2 \int_{0-}^{\infty} f^2\, dW(f) < 0.$$

The convergence of these integrals would seem to be the only assumption needed to justify the result.

Proceeding from (25), a kinematic consideration provides a short cut to the alarm rate. We choose to think of (25) as giving the position and velocity distribution of a gas in the plane; and we desire the rate at which matter diffuses out of the circle of radius V about 0. Consider the rate of diffusion over an element h of arc of this circle. The rotational invariance of the distribution makes this rate independent of where on the circle h is placed, and so we place it at $z = V$. We may think of it as a small vertical segment[12] whose endpoints are V, $V + ih$. The particles that will cross with velocity w in one second are those[12] in the parallelogram of area uh, determined by h and the vector $w = u + iv$, provided $u > 0$, and hence have relative mass $F(V, w)uh$. Integrating overall relevant (u, v), we obtain

$$h \int_{-\infty}^{\infty} \int_{0}^{\infty} F(V, u + iv)u\, du\, dv. \quad (27)$$

[12] To a suitable degree of approximation, of course.

his is the rate of diffusion of mass over an arc of length h f the circle. Multiplying by $2\pi V/h$ yields the diffusion ate for the whole circle. Laying aside the kinematic terinology which has now served its heuristic purpose, we ay that this [*i.e.*, (27) with $h = 2\pi V$] is the probability[12] hat a particle will escape from the circle per unit of time; nd this is, of course, the *alarm rate*.

It remains to evaluate (27). We set $z = V$, $w = u + iv$, n (25), and the *alarm rate* is readily evaluated as

$$Vr^{-1} \exp(-2^{-1}r^{-1}V^2)[(D/2\pi r)^{\frac{1}{2}} \exp(-2^{-1}r^{-1}V^2Rr_1) + Vr^{-1}\mathbf{R}r_1 \operatorname{erf}(VRr_1(Dr)^{\frac{1}{2}})] \qquad (28)$$

here

$$\operatorname{erf}(y) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{y} \exp(-2^{-1}x^2)\, dx.$$

In the stationary case, there is considerable simplifiation because the real part $\mathbf{R}r_1$ of $r_1 = R_c'(0)$ is 0, and e have nothing but a constant, $(D/2\pi r)\frac{1}{2}$, multiplied by e envelope's probability density at V [see (9)]. This is ice's result.[13]

VI. Mobile Filters

The generality of (28) in going beyond the stationary ase might have been pointless were it not for the fact at "Gaussian noise" may be encountered which is not ationary. Of course in that case a pre-envelope process nnot be constructed via (7). A consideration of the nesis of such noise is helpful, however. It usually arises ecause stationary noise is sent through a mobile (= nonstationary) device. It often happens that a formula of the type

$$\int_{-\infty}^{\infty} A_t(s)x(s)\, ds = z(t) \qquad (29)$$

can be discovered such that for a real input $x(t)$, $z(t)$ represents the pre-envelope for the output, *i.e.*, $z(t)$ is the real output[14] and $|z(t)|$ is the envelope of the output. Then the function (13) for the pre-envelope of the output is as follows

$$\begin{aligned} r(t, s) &= \tfrac{1}{2}E\left[\iint x(u)x(v)\overline{A_t(u)}A_s(v)\, du\, dv\right] \\ &= \tfrac{1}{2}\iint R(u - v)A_s(v)\overline{A_t(u)}\, dv\, du. \qquad (30) \end{aligned}$$

If A_s has a Fourier transform

$$\phi_s(f) = \int A_s(v) \exp(-2\pi i f v)\, dv$$

then the Fourier inversion formula enables us to write also

$$r(t, s) = \tfrac{1}{2}\int \phi_s(f)\overline{\phi_t(f)}\, dW(f). \qquad (31)$$

Acknowledgment

Thanks are due T. L. Gottier and Dr. E. Ackerlind of the Radio Corporation of America, Los Angeles, Calif., for encouragement and to the second named especially for enlightening discussion.

[13] Rice, *op. cit.* See (4.8).

[14] Often (29) is more elegant than its real part.

Correspondence

A Question of Terminology

An important concept in information theory is the amount of information reaching a receiver about a message sent by a transmitter. Shannon introduced[1] this concept as the rate of actual transmission (over a noisy discrete channel); *i.e.*, the source rate minus the equivocation, $H(x) - H(x)$.

Subsequently, Woodward[2] has spoken of "information transfer" and Fano[3] of "mutual information." Various other suggestions have been made informally, among them "transinformation" (Kretzmer) and "co-information" (Slepian).

It may be worth noting that such terms as "mutual information" or "co-information" underline the mathematical symmetry (with respect to transmitted and received symbols) demonstrated by Shannon.[1] "Information transfer" emphasizes the physical process; *i.e.*, the flow of information from transmitter to receiver. "Transinformation" still hints at this, while presumably being appropriate from the mathematical viewpoint as well. It is felt that a simple, descriptive term is needed; it would be of interest to Committee 11 to learn the opinions of the PGIT membership—particularly on the term "transinformation," because it is now working on a list of standard definitions including this term.

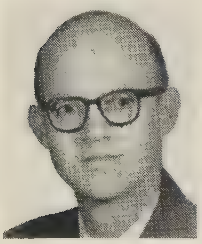
J. G. Kreer, *Chairman*
Information Theory and
Modulation Systems Committee
IRE Technical Committee 11

[1] C. E. Shannon, "A mathematical theory of communication," *Bell Sys. Tech. J.*, vol. 27, pp. 379–423, July and pp. 623–656; October, 1948.

[2] P. A. Woodward, "Probability and Information Theory, with Applications to Radar," McGraw-Hill Book Co., Inc., New York, N.Y.; 1953.

[3] R. M. Fano, unpublished notes.

Contributors



R. Arens

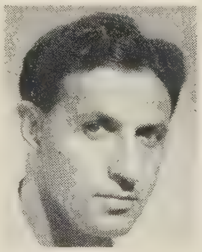
Richard Arens was born in Iserlohn, Germany on April 24, 1919. He received the B.A. degree in 1941 from the University of California, Los Angeles, in mathematics and physics. He obtained the A.M. degree in 1942 and the Ph.D. degree in 1945, in mathematics from Harvard University. After two years at the Institute for Advanced Study, he joined U.C.L.A., where he is presently professor of mathematics. His main interest lies in topological algebra.

Dr. Arens is a consultant to the Radio Corporation of America in Los Angeles.

R. F. Baum

R. F. Baum was born in Czechoslovakia on August 18, 1911. He obtained the M.S. degree in electrical engineering at the Deutche Technische Hochschule at Prague and the M.S. degree in radio engineering at the École Superieure d'Electricité at Paris.

He worked at the Federal Telecommunication Laboratories on the development of uhf direction finders from 1943 to 1945 and of Tacan from 1950 to 1953. From 1945 to 1949 he was engaged with research and the development of microwave links at the Raytheon Manufacturing Company. During 1949–1950 he was doing systems work on electronic fuzes at the National Bureau of Standards. Currently he is employed as staff engineer at the W. L. Maxson Corporation, being mainly concerned with the analysis of radar and ecm systems.



R. Bellman

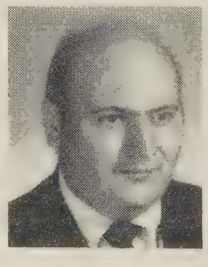
Richard Bellman was born in New York in 1920. He received the B.A. degree from Brooklyn College in 1941, the M.A. from the University of Wisconsin in 1943, and the Ph.D. from Princeton University in mathematics in 1946. At Princeton he was successively a Fine instructor and an assistant professor until 1948 when he accepted an associate professorship at Stanford University. In 1952, he left Stanford to join The RAND Corporation.

In 1942 and 1943, he taught electronics in the U.S. Air Force preradar schools at Scott Field, Ill., and Truax Field, Wis. In 1944 he was a mathematical physicist at the U.S. Naval Radio and Sound Laboratory at Point Loma, San Diego, Calif. From December, 1944, until March, 1946, he was a member of the Special Engineering Division, U.S. Army, at Los Alamos, N.M.

He was associated, in 1951, with the Matterhorn Project at Princeton. He was visiting Professor of Engineering at UCLA in 1956 and at the present time is a consultant for the AEC, for Booz, Allen and Hamilton, management consultants, and for the Willow Run Laboratories, University of Michigan.

Dr. Bellman is the author of two books, "Stability Theory" and "Dynamic Programming," and of over 150 research papers.

He is a member of the Council of the American Mathematical Society and of the Committee on Applied Mathematics.



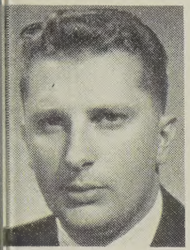
M. Blum

Marvin Blum (M'56) was born on June 18, 1928 in New York, N.Y. He received the B.S. degree from Brooklyn College in June 1948, and has taken graduate courses in mathematics, physics, and electrical engineering from George Washington University, American University, Maryland University, National Bureau of Standards School and U.C.L.A. Extension.

Mr. Blum worked at the National Bureau of Standards in the Central Radio Propagation Laboratory until 1950. He then transferred to the Ordnance Division, where he conducted radar reflection studies relating to missile proximity fuzes.

Since July, 1954, Mr. Blum has been employed at Convair, San Diego, Calif., where he is conducting theoretical investigations in smoothing and prediction filters, noise simulation, and data reduction. He is presently working in the newly organized Convair Astronautics division.

Robert Kalaba was born in Mount [V]ernon, N.Y., on September 21, 1926. He [re]ceived the A.B. from New York University where he is a candidate for the Ph.D. in mathematics.



R. KALABA

He has served as a teaching Fellow at N.Y.U. and since 1951 has been a mathematician in the electronics department of the RAND Corporation, specializing in communications problems. He is a lecturer in engi[n]eering at U. C. L. A.

He is a member of Phi Beta Kappa and [th]e American Mathematical Society.

❖

R. A. Silverman (M '54) was born on June [1]9, 1926, in Boston, Mass. He received the [A].B. degree from Harvard University in 1946, the M.A. degree from Columbia University in 1948, and the Ph.D. degree from Harvard in 1951.

[R]. A. SILVERMAN

For three years Dr. Silverman was associated with the Lincoln Laboratory at Massachusetts Institute of Technology, and was also a research associate in the Department of Electrical Engineering.

Dr. Silverman is currently a research associate at the New York University Institute of Mathematical Sciences in the Division of Electromagnetic Research.

He is a member of Phi Beta Kappa, Sigma Xi, and the American Physical Society.

❖

Peter Swerling (M'56) was born in New York, N.Y., on March 4, 1929. He received the B.S. degree in mathematics from California Institute of Technology in 1947, the B.A. degree in economics from Cornell University in 1949, and the M.A. and Ph.D. degrees in mathematics from the University of California at Los Angeles in 1951 and 1955.

PETER SWERLING

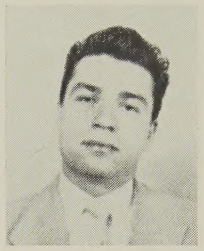
Dr. Swerling worked on Project Rand at Douglas Aircraft Co. in 1947 and 1948. From September, 1949 to January, 1950, he was teaching assistant in mathematics at U.C.L.A. In 1949 he was employed by RAND Corporation, working full time there in 1952. From 1956-1957 he served as visiting research assistant professor at the Control Systems Laboratory of the University of Illinois, returning then to RAND Corporation where he is presently employed.

Theory of random noise, especially as applied to radar performance, has been Dr. Swerling's major field of research.

He is a member of Phi Beta Kappa, Sigma Xi, Pi Mu Epsilon, American Mathematical Society, and the Society for Industrial and Applied Mathematics.

❖

D. Youla was born in Brooklyn, N. Y., on October 17, 1925. He received the B.E.E. degree from City College in January, 1947, and the M.S. degree from New York University in June, 1950.



D. YOULA

From 1947 to 1949 he was employed as an instructor in the Department of Electrical Engineering at C.C.N.Y. He attended the N.Y.U. Graduate School of Mathematics as a full-time student from 1948-1950, and for the next two years was at Fort Monmouth and Brooklyn Naval Shipyard working on problems of uhf and microphonics. In 1952 he joined the communication group at the Jet Propulsion Laboratories, Pasadena, Calif., and participated in the design of antijam radio links for guided missiles. In 1954 he began his present association with Polytechnic Microwave Institute where he is engaged in the practical and theoretical study of codes for combating noise and improving efficiency.

INFORMATION FOR AUTHORS

Authors are requested to submit editorial correspondence or technical manuscripts to the Publications Chairman for possible publication in the PGIT Transactions. Papers submitted should include a statement as to whether the material has been copyrighted, previously published, or accepted for publication elsewhere.

Papers should be written concisely, keeping to a minimum all introductory and historical material. It is seldom necessary to reproduce in their entirety previously published derivations, where a statement of results, with adequate references, will suffice.

To expedite reviewing procedures, it is requested that authors submit the original and two legible copies of all written and illustrative material. The manuscript should be double-spaced, and the illustrations drawn in India ink on drawing paper or drafting cloth. Each paper should include a carefully written abstract of not more than 200 words. Upon acceptance, papers should be prepared for publication in a manner similar to those intended for the Proceedings of the IRE. Further instructions may be obtained from the Publications Chairman. Material not accepted for publication will be returned.

IRE Transactions on Information Theory is published four times a year, in March, June, September, and December. A minimum of one month must be allowed for review and correction of all accepted manuscripts. In addition, a period of approximately two months is required for the mechanical phases of publication and printing. Therefore, all manuscripts must be submitted three months prior to the respective publication dates. In addition, the IRE National Convention Record is published in July, and the IRE WESCON Convention Record in the Fall. A bound collection of Information Theory papers delivered at these conventions is mailed gratis to all PGIT members.

All technical manuscripts and editorial correspondence should be addressed to Laurin G. Fischer, Federal Telecommunication Labs., 492 River Road, Nutley, N. J. Local Chapter activities and announcements, as well as other nontechnical news items, should be addressed to Nathan Marchand, Marchand Electronic Labs., Riversville Road, Greenwich, Conn.